

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
14 March 2002 (14.03.2002)

PCT

(10) International Publication Number  
WO 02/20837 A2

(51) International Patent Classification<sup>7</sup>: C12Q 1/68

Nader [SE/US]; Stanford DNA Sequencing & Technology Center, 855 California Avenue, Palo Alto, CA 94304 (US).

(21) International Application Number: PCT/GB01/04042

(74) Agents: GARDNER, Rebecca et al.; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

(22) International Filing Date:  
10 September 2001 (10.09.2001)

(81) Designated States (national): AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EC, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
0022069.9 8 September 2000 (08.09.2000) GB

(71) Applicants (for all designated States except US): PY-ROSEQUENCING AB [SE/SE]; Vallongatan 1, S-752 28 Uppsala (SE). THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY [US/US]; Suite 350, 900 Welch Road, Palo Alto, CA 94304-1850 (US).

(71) Applicant (for GB only): GARDNER, Rebecca [GB/GB]; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): RONAGHI, Mostafa [SE/US]; Stanford DNA Sequencing & Technology Center, 855 California Avenue, Palo Alto, CA 94304 (US). EKSTRÖM, Björn [SE/SE]; PyroSequencing AB, Vallongatan 1ala, S-752 28 Uppsala (SE). POURMAND,

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/20837 A2

(54) Title: METHOD

(57) Abstract: The method of the invention relates to a method of typing one or more nucleic acid molecules, said method comprising: simultaneously or sequentially performing two or more primer extension reactions, each primer binding at a different predetermined site in said nucleic acid molecule(s), and determining the pattern of nucleotide incorporation to obtain a test pattern for said nucleic acid molecule(s) which is optionally compared with one or more reference patterns to type the said nucleic acid molecule(s).

Method

This invention relates to a method of typing using  
5 nucleic acid, and in particular to an improved method of  
genotyping.

Typing, e.g. genotyping, can be particularly  
advantageous for medical diagnosis, prognosis and  
treatment. For example, identification of the microbe  
10 responsible for infection allows correct treatment to be  
administered. It has been shown that microbes may now  
readily be identified by typing (i.e. by identifying  
genomic signature patterns characteristic of a  
particular microbe). Typing one or more variable  
15 regions in a gene or genes or other nucleic acid  
sequence of an individual can reveal markers of  
predisposition to a particular disease, condition or  
syndrome, and may also point to the best method of  
treatment of the foregoing. Typing methods are also  
20 useful for genomic analyses (e.g. in typing  
polymorphisms or allelic variations), tissue typing or  
environmental monitoring and contamination testing etc.

Conventional assays for detection of bacterial or  
viral species, or for detecting mutations or  
25 polymorphisms in a DNA sequence include using the  
polymerase chain reaction (PCR) method. This method is  
designed to permit selective amplification of a  
particular target DNA sequence or sequences, determined  
by the nature of the amplification primers used. To  
30 permit such selective amplification, some prior  
knowledge of the sequence of the DNA is required,  
enabling the construction of two oligonucleotide primer  
sequences, known as amplimers. One amplimer hybridises  
at or towards the 5' end of one of the strands of the  
35 target DNA and the other amplimer at or towards the 5'  
end of the second strand. In the presence of a DNA  
polymerase and DNA precursors (i.e. dATP, dCTP, dGTP and

- 2 -

dTTP) the primers can initiate the synthesis of new DNA strands which are complementary to the individual strands of the target DNA segment. The use of a heat-stable polymerase enables the procedure to be readily 5 repeated or cycled. The newly synthesized DNA strands act as templates for further DNA synthesis in subsequent cycles. The reaction mixture is subjected to a temperature of about 90°C to separate the double stranded DNA formed by the polymerase. The reaction 10 temperature is reduced to about 50°C to 70° to allow the single stranded DNA to anneal to the primers, and another round of DNA synthesis is performed. The DNA synthesized extends between the termini of the two primers. Preferably, the DNA polymerase used is 15 thermophilic i.e. Tag polymerase. After 30 cycles of DNA synthesis, the products of PCR will include about  $10^5$  copies of the specific target sequence. A typical PCR reaction cycle is therefore: synthesis of the separate strands by primer extension, separation of strands, 20 primer annealing, synthesis of new strands.

The chain reaction can therefore be perpetuated merely by raising and lowering the temperature.

Typing (e.g. genotyping) be performed using PCR-based techniques, for example using allele-specific 25 primers (Okamoto et al. 1992, Journal of General Virology, 73, 673-679; Widell, A et al. 1994, Journal of Medical Virology, 44, 272-279).

Currently, multiplex PCR may be used to screen 30 samples of nucleic acids for a given panel of mutations/variations within the nucleic acid sequence. This method is still cumbersome for routine diagnostics, as the use of gel electrophoresis is essential. Alternative methods rely on the use of labelled 35 nucleotides or primers, and may require complex detection strategies or mechanisms. There is thus a need for a typing method which may analyse nucleic acids with respect to 2 or more variable regions or positions

- 3 -

typically without the need for gel electrophoresis and preferably without the use of labelled nucleotides or primers.

Other methods for typing include serologically-based detection methods (Viazov, et al. 1994, Journal of Virological Methods, 48, 81-91; Schroter, M., 1999, Journal of Medical Virology, 57, 230-234), line probe assay (Stuyver, L., 1993, Journal of General Virology, 74, 1093-1102, Stuyver, L., 1996, Transfusion, 36, 552-558), and restriction fragment length polymorphism (McOmish et al. 1993, Transfusion, 33, 7-13; Buoro, S. 1999, Intervirology, 42, 1-8) are well known in the art. However, sequencing continues to be regarded in the art as the "gold standard" method for typing. Accordingly, a sequencing-based typing method which avoids the drawbacks mentioned above would represent a considerable advance in the art.

PCR is also commonly used in the detection of microbes, i.e. bacteria or viruses. However, conventional PCR assays are limited in diagnostic applications, and generally only indicate whether or not a microbe, or a particular sequence is present. For many infections, e.g. viral infections such as hepatitis C viral infection, the infecting microorganism may occur in a number of different sub-types, for example at least seven subtypes (or genotypes) are known of the HCV virus. It would be advantageous in such circumstances not only to determine that the general "class" (or genus or species) of infecting microorganism is present (e.g. HCV virus), but also to determine which of the sub-types is present.

Similarly, genomic studies have now revealed that many other diseases or disorders may be associated with genetic variations (e.g. mutations, allelic variations or polymorphisms (e.g. single nucleotide polymorphisms, (SNPs))), and that the presence of such variations may indicate a risk or predisposition to a disease or

disorder, or may even indicate or predict how an individual may respond to a particular treatment for that disease or disorder (this latter effect is referred to as "pharmacogenomics"). Accordingly, in clinical 5 science, the analysis (typing) of such variations may be of importance.

Microbial subtypes and clinically informative polymorphisms (or other genetic variations) frequently are characterised by combinations of genetic variations 10 (i.e. variations in multiple (i.e. two or more) positions or regions of the genome etc.). Accordingly, for typing purposes in such situations it is necessary to "type" (or identify) more than one variation 15 (polymorphism). In other words, it is necessary to type (or study or identify) a polymorphic pattern (a pattern of genetic variations) which pattern may cover more than one region of the genome to be studied. (The term "polymorphic pattern" is used herein broadly to include patterns, or combinations, of two or more (e.g. 3, 4, 5, 20 6, 7, 8, 9, 10 or more) of any type of genetic variation, e.g. mutations, allelic variants, polymorphisms of any type etc.). It will be understood that the variation can be an insertion or deletion of one or more nucleic acid residues.

25 Thus, for many microorganisms, diseases or predisposition to disease, an identification of the exact type of microbe, or genetic variation, present is needed to make a proper diagnosis or prognosis and in order to achieve this it is necessary to study more than 30 one genetically variant position or region. As mentioned above, PCR is an extremely useful tool for the amplification and/or identification of a specific sequence of DNA, but to use conventional PCR techniques to determine the genotype of a nucleic acid molecule 35 based on multiple genetic variations requires a repeated and multiple number of individual reactions to be performed, which would be cumbersome, time-consuming and

- 5 -

expensive to perform using conventional technologies and procedures such as e.g. electrophoresis or labelling technologies. There is therefore a need for a typing assay that is accurate and reliable, has a short 5 analysis time and is quick and easy to perform. The present invention addresses this need.

In particular, it has now been found that a simple, reliable, and accurate method for obtaining typing (sequence) information about a plurality of variable 10 sites within target nucleic acid, may be performed using a primer extension reaction system using two or more specific primers designed to bind at or near to these variable sites, allowing primer extension reactions to be carried out on each primer annealed to a template 15 nucleic acid sequence, either sequentially or simultaneously, and detecting the pattern of nucleotide incorporation in said primer extension reactions. The pattern of nucleotide incorporation providing the typing information about said variable sites.

20 This new method of the invention thus combines a multiplexing approach (i.e. an approach relying on the simultaneous or parallel performance of multiple reactions) with a particular strategy for detecting the result of the multiple primer extensions, namely 25 detecting the pattern of nucleotide incorporation.

The method is particularly suited to automation e.g. in systems where reaction and reagent dispensing steps take place in a microtitre plate format. The methods are particularly suitable for identifying 30 microbial species and subtypes thereof, but may also find application in other typing procedures e.g. typing of polymorphisms, e.g. for tissue typing or in clinical applications.

As described further below the present invention is 35 advantageously based on a method of "sequencing-by-synthesis" (see e.g. US-A-4,863,849 of Melamede). This is a term used in the art to define sequencing methods

which rely on the detection of nucleotide incorporation during a primer-directed polymerase extension reaction. The four different nucleotides (i.e. A, G, T or C nucleotides) are added cyclically or sequentially 5 (conveniently in a known order), and the event of incorporation can be detected in various ways, directly or indirectly. This detection reveals which nucleotide has been incorporated, and hence sequencing information; when the nucleotide (base) which forms a pair (according 10 to the normal rules of base pairing, A-T and C-G) with the next base in the template target sequence is added, it will be incorporated into the growing complementary strand (i.e. the extended primer) by the polymerase, and this incorporation will trigger a detectable signal, the 15 nature of which depending upon the detection strategy selected.

Accordingly, the present invention provides a method of typing 1 or more nucleic acid molecules, said method comprising:

20 simultaneously or sequentially performing two or more primer extension reactions, each primer binding at a different predetermined site in said nucleic acid molecule(s), and determining the pattern of nucleotide incorporation to obtain a test pattern (or 25 "fingerprint") for said nucleic acid molecule which is optionally compared with one or more reference patterns to type the said nucleic acid molecule(s).

30 Preferably, the primer extension reactions occur simultaneously, i.e. both or all primers are annealed and are capable of primer extension at the same time. It will, of course, be appreciated that each individual primer can only be extended if a nucleotide is added to the reaction mix which is complementary to the next nucleotide in the template. Thus for each nucleotide 35 addition, not every primer (or even any primer) will actually be extended and the term 'simultaneous' must be interpreted with this in mind.

The method of the invention may be used to type a nucleic acid molecule containing two or more sites at which its sequence may be variable ("variable sites") and each said primer binds at a site lying at or near to 5 a variable site. Different nucleotides may be added sequentially to perform the primer extension reactions, and are described further below.

Alternatively, the method of the invention may be used to type two or more nucleic acid molecules 10 containing 1 or more sites at which the sequence may be variable ("variable sites") and each primer binds at a site lying at or near to a variable site. Different nucleotides may be added sequentially to perform the primer extension reactions, and are described further 15 below. This embodiment may be particularly useful when it is desired to obtain information about variable sites within the related genes, for example SNPs in Factor V Leiden and Prothrombin (FII) which are genetic risk factors for developing venous thrombosis.

20 The term "typing" as used herein includes any method of analysing the nucleotide sequence of the nucleic acid molecule to be analysed (i.e. the "test" or target nucleic acid). More particularly, the typing method of the invention includes methods for detecting, 25 identifying or analysing genetic or sequence variation (e.g. genomic variation) in a target nucleic acid molecule or molecules (as mentioned above, this may be e.g. mutation, allelic variation, polymorphisms etc.). Methods of the invention thus include methods of identifying, differentiating or distinguishing a nucleic acid molecule or molecules. Since the typing method of 30 the invention relies on detecting genetic variation in a nucleic acid molecule or molecules, it may be regarded as a method of genotyping. It will be understood that a nucleic acid molecule may itself be typed and also that 35 a given variable site within a nucleic acid molecule may be typed.

- 8 -

"Genotyping" according to the present invention thus involves determining the genotype of the target nucleic acid molecule(s). In the context of this specification, the "genotype" may be regarded as the 5 particular combination or pattern of the genetic variations which are studied or analysed in the method of the invention, which is exhibited (or expressed) by the nucleic acid molecule(s) in question. The genotype may thus comprise the combination (or pattern) of 10 particular alleles (i.e. variations) which are found at the particular loci investigated.

In other words, the genotype is a combination or pattern of multiple genetic variations (or "variable sites") in target nucleic acid. The genetic variations 15 which comprise or make up the genotype may be those selected for study in the method of the invention (notwithstanding that other genetic variations may also be present in the molecule, which are not investigated). As mentioned above, "multiple" as used herein means 2 or 20 more (or 3,4,5,6,7,8,9,10 or more), and the genetic variations (or "variable sites") may be polymorphisms (e.g. SNPs), insertions, deletions, mutations, hypervariable regions, variable motifs, or allelic variations, etc. According to the methods of the 25 invention, 2 or more, preferably 3 or more, e.g. 3-7 variable sites are investigated simultaneously. Unless two of such variable sites are found close together, e.g. with 50 nucleotides, preferably within 30 nucleotides, preferably within 20 nucleotides a separate 30 primer will be required in order to type each variable site. So each primer will be responsible for generating typing information about one or more variable sites, a primer will therefore in effect have its 'own' variable site(s).

Conveniently, the target nucleic acid may be DNA, although typing of RNA (e.g. mRNA) is also within the scope of the invention. If it is desired to type a RNA

sample, the method may additionally include the step of generating cDNA from the RNA template, conveniently by using reverse transcriptase. Alternatively, if desired, the primer extension reactions may be performed directly 5 on the RNA template.

The target nucleic acid may thus be any nucleic acid, isolated or synthetic, in any desired or convenient form. It may thus be genomic DNA, or 10 isolated mRNA which may be used directly for analysis by the method of the invention, or it may be a nucleic acid product derived therefrom (or corresponding thereto), e.g. by synthesis, such as cDNA as mentioned above, or an amplification product (e.g. PCR amplicon), clones or library products etc.

15 The nucleic acid molecule(s) may be obtained or derived from any convenient source, which may be any material containing nucleic acid, and all biological and clinical samples are included as possible sources i.e. any cell or tissue samples of an organism, or any body 20 fluid or preparation derived therefrom, as well as cell cultures, cell preparations, cell lysates etc. Environmental samples e.g. soil and water samples or food samples are also included. The samples may be freshly prepared or they may be prior-treated in any 25 convenient way e.g. for storage.

Representative sources of nucleic acid thus 30 include, for example, foods and allied products, clinical and environmental samples. However, the source will generally be a biological sample, which may contain any viral or cellular material, including all prokaryotic or eukaryotic cells, viruses, 35 bacteriophages, mycoplasmas, protoplasts and organelles. Such biological material may thus comprise all types of mammalian and non-mammalian animal cells, plant cells, algae including blue-green algae, fungi, bacteria, protozoa etc. Representative sources thus include whole blood and blood-derived products such as plasma, serum

- 10 -

and buffy coat, urine, faeces, cerebrospinal fluid or any other body fluids, tissues, cell cultures, cell suspensions etc.

The nucleic acid may be provided for investigation 5 in any convenient form and conveniently will be contained in a sample, e.g. an aqueous sample (e.g. in a buffer etc.). The nucleic acid may be prepared for the typing method, as desired, according to techniques well known in the art, e.g. isolation, purification, cloning, 10 copying, amplification, etc.

In carrying out the method of the invention, two or 15 more primers ("extension primers") are provided which bind to the target nucleic acid at a predetermined site, each primer binding site being different, so that multiple different primer extension reactions are performed. The extension primers are designed or selected so that their extension products overlap (or comprise) a site (e.g. locus or region) of sequence variability (i.e. genetic variation) in the target 20 nucleic acid. In other words, the primers bind to the target nucleic acid, or near to (e.g. within 1 to 40, 1 to 20, 1 to 10, or 1 to 6 bases of), a variable site. As mentioned above, such variable sites constitute the genotype of the target nucleic acid.

At least two extension primers are required to 25 carry out the method, preferably at least three. However, the number of primers may be varied according to choice, for example, depending on the complexity of the system under study, and the detail of the 30 information it is desired to obtain. Thus, for example, 3, 4, 5, or 6, or more extension primers (e.g. 3 to 15, or 3-10) may be used.

Thus, the term "variable site" refers to a site 35 (e.g. locus or region) of a nucleic acid molecule which can differ in different genotypes. As defined above, the variable site may be a polymorphism or motif etc. Nucleic acid markers used for typing normally contain

- 11 -

both conserved/semi-conserved and variable regions. Thus, each "type" will comprise a region of sequence variation, wherein this region (i.e. the sequence, or base identity, at that site) can be different from other 5 types. In the method of the invention, at least two potential variable sites are examined, and, when one target nucleic acid molecule is typed, said nucleic acid molecule thus contains 2 or more (i.e. multiple) variable sites. Where 2 or more target nucleic acid 10 molecules are typed, said nucleic acid molecules thus each contain 1 or more variable sites.

It will be understood by the skilled person in the art that any desired combination of variable sites can be analysed by the method of the invention. The 15 variable sites do not have to be restricted to a single gene, coding region, non-coding region or nucleic acid molecule, but may be found anywhere in the target genome. It will further be understood that the variable site can be of any length, optionally 1 to 20 nucleotides, preferably 1 to 10 nucleotides in length. Typically, however, the variable site may comprise only 20 a single or a few (e.g. 1-6, e.g. 1, 2, 3, 4, 5 or 6) nucleotides at which the sequence of the target nucleic acid may be variable. Thus, for example, a virus such 25 as hepatitis C virus (HCV) may contain regions which are conserved between sub-types, but which nonetheless contain sites which may vary between subtypes. Such variable sites (which may typically be 1 to 3 nucleotides in length) may thus be used to distinguish 30 between the various subtypes. In HCV such a conserved region containing variable sites is the 5' untranslated region (5'UTR), and this may conveniently be used in a genotyping assay method of the invention, as described further in Example 1 below.

35 Other microorganisms will analogously have similar such regions in their genomes, containing variable sites, which may similarly be used in the method of the

- 12 -

invention. For other typing applications e.g. typing of polymorphisms, regions of sequence variability, analogously containing polymorphic sites, may similarly be identified. For example, SNPs in the Renin-  
5 Angiotensinogen-Aldosterone system (RAAS) may be assessed using primers position in conserved regions of the genes. The primer can be position at or near to the SNP site. Figure 5 shows the positioning of three different extension primers, wherein the 3' end of the  
10 primer is 4 bases, 5 bases or 10 bases from the SNP position. The SNP is EU6 (ACE T3409C).

It will be understood that in order to perform the invention the primer binding sites should be available in all possible variants (genotypes) of the nucleic acid  
15 molecule(s) under study. Such primer binding sites will therefore advantageously lie in regions which are common to, or substantially conserved between, the different variants. This may readily be achieved by selecting the primer binding sites to lie in conserved/semi-conserved  
20 regions as discussed above.

The primer extension reactions conveniently may be performed by sequentially adding the nucleotides to the reaction mixture (i.e. a polymerase, and primer/template mixture). Advantageously the different nucleotides are  
25 added in known order, and preferably in a pre-determined order. In a convenient embodiment of the invention described in Example 1 below, the 4 different nucleotides (i.e. A, G, T and C nucleotides) are added sequentially in a predetermined order of addition. It thus forms a preferred aspect of the invention that the  
30 nucleotides are added sequentially in a predetermined order of addition. Therefore, the order of addition can be tailored to the nucleic acid(s) to be typed and the primers used. It will therefore be seen that the order of addition will not necessarily be cyclical e.g.  
35 A T G C A T G C but can be e.g. C G C T A G A.

As each nucleotide is added, it may be determined

- 13 -

whether or not nucleotide incorporation takes place.

Advantageously, as described in more detail below, it may further be determined the amount (i.e. how many) of each nucleotide incorporated. In this manner, the 5 pattern of nucleotide incorporation may be determined. In other words, the step of determining the pattern of nucleotide incorporation may comprise determining (or detecting) whether or not, and which, nucleotide is incorporated. Advantageously, this step also includes 10 determining the amount of each nucleotide incorporated. Such a quantitative embodiment, wherein nucleotide incorporation is determined quantitatively, represents a preferred aspect of the invention.

In this manner, a "pattern" or "fingerprint" may be 15 obtained for the target nucleic acid. This pattern comprises the base identity (i.e. sequence) of the particular variable sites identified for that nucleic acid molecule. In other words, the pattern corresponds to the genotype of the target nucleic acid. The 20 genotype may thus readily be identified by comparing the pattern obtained to a reference pattern (or a "standard pattern"), or a panel of reference patterns (i.e. one or more, e.g. two or more e.g. 1 to 20, 1 to 15, 1 to 10, 1 to 6 or 1 to 3). A reference pattern may readily be 25 obtained by determining the pattern of nucleotide incorporation using the extension primers in question on reference nucleic acid molecules of known genotype (e.g. a known microbial subtype or a known polymorphic pattern).

30 Alternatively, the 'reference pattern' can be theoretically derived from knowledge of the variable sites, as shown in the later Examples. It may then not be necessary actually to compare the pattern obtained with a reference pattern, the desired typing/sequence 35 information can be read from the pattern obtained. Once the extension primers for each variable site have been selected and the order of addition of nucleotides

determined, it is possible to determine a theoretical output from a primer extension reaction. Figure 6 shows the theoretical output from sequencing two variable sites individually, and the combination of extending 5 both extension primers simultaneously. The theoretical reference pattern is shown for 2 variable sites present as heterozygotes. The primers used bound 3' to the sequences shown.

Thus, by identifying (or recognising) the pattern 10 obtained for a target nucleic acid molecule, the genotype of the molecule may be identified (or recognised). Conveniently, test patterns and reference patterns may be compared using pattern recognition software.

15 In order to perform the invention, it may be advantageous or convenient first to amplify the nucleic acid molecule by any suitable amplification method known in the art. The target nucleic acid would then be an amplicon. Suitable in vitro amplification techniques 20 include any process which amplifies the nucleic acid present in the reaction under the direction of appropriate primers. The amplicon method may thus preferably be PCR, or any of the various modifications thereof e.g. the use of nested primers, although it is 25 not limited to this method. Those skilled in the art will appreciate that other amplification procedures may also be used, such as Self-sustained Sequence Replication (3SR), NASBA, the Q-beta replicase amplification system and Ligase chain reaction (LCR) 30 (see for example Abramson and Myers (1993) Current Opinion in Biotech., 4: 41-47).

If PCR is used to amplify the nucleic acid, suitable primers, as discussed previously, are designed 35 to ensure that the region of interest within the nucleic acid sequence (i.e. the region containing the variable sites), is amplified. PCR can also be used for indiscriminate amplification of all DNA sequences,

- 15 -

allowing amplification of essentially all sequences within the sample for study (i.e. total DNA). Linker-primer PCR is particularly suitable for indiscriminate amplification, and uses double stranded oligonucleotide 5 linkers with a suitable overhanging end, which are ligated to the ends of target DNA fragments. Amplification is then conducted using oligonucleotide primers which are specific for the linker sequences. Alternatively, completely random oligonucleotide primers 10 may be used in conjunction with DOP-PCR (degenerate oligonucleotide-primed) to amplify all the DNA within a sample. If the variant sites to be typed by the method of the invention are present in discrete areas of the genome, multiplex PCR can be used to amplify nucleic 15 acid sequences from the genome containing the variable sites. Therefore, multiple fragments can be amplified in a single PCR reaction.

In the method of the invention, several sequences may need to be amplified, to allow several regions (e.g. 20 containing different variable sites) to be analysed. Therefore, several appropriate amplification primers may need to be synthesized to allow the selective amplification of several sequences in the target nucleic acid. It will therefore be understood that a number of 25 different nucleic acid molecules may be present in the reaction mixture.

One or more of the amplification primers used in the amplification reaction, may be subsequently used as an "extension primer", but this will preferably be a 30 different primer.

It will be appreciated that the sequence and length of the oligonucleotide amplification and extension primers to be used in the amplification and extension steps, respectively, will depend on the sequence of the 35 target nucleic acid, the desired length of amplification or extension product, the further functions of the primer (i.e. for immobilization) and the method used for

- 16 -

amplification and/or extension. Appropriate primers may readily be designed applying principles and techniques well known in the art.

Advantageously, as mentioned above, extension  
5 primers will bind near (e.g. within 1-40, 1-20, 1-10 or  
1-6, preferably within 1-3 bases), substantially  
adjacent or exactly adjacent to the variable site of the  
target nucleic acid and will be complementary to a  
conserved or semi-conserved region of the nucleic acid.  
10 In certain embodiments, as described in Example 1 for  
instance, all primers will bind substantially adjacent  
to variable sites within the target nucleic acid (i.e.  
adjacent or within 3 bases of the variable site). In  
other embodiments, see for example Example 3, the  
15 primers will be staggered so that one is very close to  
its variable site, another is some distance away, e.g.  
4-10 nucleotides distant and a third primer is 7 or more  
e.g. 8-16 nucleotides distant from its (first) variable  
site. Figure 7 depicts this principle.

20 In order for the method of the invention to be  
performed, knowledge of the sequence of the conserved or  
semi-conserved region is required in order to design an  
appropriate complementary extension primer. An  
extension primer is provided for each of the variable  
25 regions, each being specific for a site at or near to  
the variable site. The specificity is achieved by  
virtue of complementary base pairing. For all  
embodiments of the invention, primer design may be based  
upon principles well known in the art. It is not  
30 necessary for the extension or amplification primer to  
have absolute complementarily to the binding site, but  
this is preferred to improve the specificity of binding.

The extension primer may be designed to bind to the  
sense or anti-sense strand of the target nucleic acid.

35 In a preferred embodiment of the invention, the  
extension primers are designed to bind to the target  
nucleic acid near to the variable sites in such a way

- 17 -

that upon the addition of nucleotides in a predetermined manner, the typing of each variable site takes place discretely. Thus analysis of a given variable site is not complicated by a positive incorporation signal from 5 other variable or conserved regions. As shown in Example 1, it is possible to interpret the test pattern and allow for signals from nucleotide incorporation at more than the primer, but preferably when one primer is extending over a variable site, the other primers will 10 be silent. Thus, if nucleotide incorporation takes place at one variable site, there is preferably no nucleotide incorporation at the other variable site(s). For example in the theoretical pattern shown in Figure 15 6, the extension primers are positioned in such a way that, upon the pre-determined sequential addition of nucleotides, each variable site is typed discretely, even though primer extension occurs simultaneously at other points - e.g. the second dispensation of nucleotide A. In this preferred embodiment, only the 20 variable site is sequenced when an extended primer reaches its variable site; the other primers are not extended as the nucleotide added to the reaction mixture is not complementary to the next base in the templates for the other primer-extension reactions. Thus, the 25 primers should be designed in parallel with the order of addition of the nucleotides. Primer design software can be used to determine the actual sequence of the primer once the 3' end has been fixed. Preferably, Pyrosequencing Primer Design Software is used.

30 Figure 7 shows a simplified set of multiplex primer extension reactions wherein the extension primers are placed at differing distances from the variant site (shown as X). This enables a pre-determined pattern of 35 nucleotide addition to be performed wherein the variant site is sequenced in isolation. As can be envisaged, variations in the nucleotide sequence upstream of the different variable sites may mean that the primers need

- 18 -

not anneal in such a staggered manner but the pattern of nucleotide addition alone may be sufficient to ensure the (extended) primers approach and/or sequence their variable sites at different times. Thus all primers may 5 anneal at similar distances from the variable sites (or where one primer extension reaction is used to type 2 variable sites, the first variable site) e.g. substantially adjacent thereto, but where the nucleotides immediately upstream of and within the 10 variable sites vary, so the order of nucleotide addition will control the order of primer extension across the variable sites.

The "primer extension" reaction according to the invention includes all forms of template-directed 15 polymerase-catalysed nucleic acid synthesis reactions. Conditions and reagents for primer extension reactions are well known in the art, and any of the standard methods, reagents and enzymes etc. may be used in this step (see e.g. Sambrook et al., (eds), Molecular 20 Cloning: a laboratory manual (1989), Cold Spring Harbor Laboratory Press). Thus, the primer extension reaction at its most basic, is carried out in the presence of primer, deoxynucleotides (dNTPs) and a suitable polymerase enzyme e.g. T7 polymerase, Klenow or 25 Sequenase Ver 2.0 (USB USA), or indeed any suitable available polymerase enzyme. As mentioned above, for an RNA template, reverse transcriptase may be used. Conditions may be selected according to choice, having regard to procedures well known in the art.

30 The primer is thus subjected to a primer-extension reaction in the presence of a nucleotide, whereby the nucleotide is only incorporated if it is complementary to the base immediately adjacent (3') to the primer position. The nucleotide may be any nucleotide capable 35 of incorporation by a polymerase enzyme into a nucleic acid chain or molecule. Thus, for example, the nucleotide may be a deoxynucleotide (dNTP,

- 19 -

deoxynucleoside triphosphate) or dideoxynucleotide (ddNTP, dideoxynucleoside triphosphate). Thus, the following nucleotides may be used in the primer-extension reaction: guanine (G), cytosine (C), thymine (T) or adenine (A) deoxy- or dideoxy-nucleotides. Therefore, the nucleotide may be dGTP (deoxyguanosine triphosphate), dCTP (deoxycytidine triphosphate), dTTP (deoxythymidine triphosphate) or dATP (deoxyadenosine triphosphate). As discussed further below, suitable analogues of dATP, and also for dCTP, dGTP and dTTP may also be used. Modified nucleotides which include an activation or detectable group, radio or fluorescently labelled nucleotide triphosphates can also be used in the primer extension step. Dideoxynucleotides may also be used in the primer-extension reaction. The term "dideoxynucleotide" as used herein includes all 2'-deoxynucleotides in which the 3' hydroxyl group is modified or absent. Dideoxynucleotides are capable of incorporation into the primer in the presence of the polymerase, but cannot enter into a subsequent polymerisation reaction, and thus function as a "chain terminator".

If the nucleotide is complementary to the target base, the primer is extended by one nucleotide, and inorganic pyrophosphate is released. As discussed further below, in a preferred method, the inorganic pyrophosphate may be detected in order to detect the incorporation of the added nucleotide. For some variable sites, the addition of one nucleotide will be sufficient to generate typing information. However, for the majority of variable sites, data for several adjacent nucleotides will be necessary. The extended primer can serve in exactly the same way in a repeated procedure to determine the next base in the variable region, thus permitting the whole variable site to be sequenced. Different nucleotides may be added sequentially, advantageously in known order, as

- 20 -

discussed above, to reveal the nucleotides which are incorporated for each extension primer. Furthermore, in the case where the variable site is homopolymeric (i.e. contains 2 or more identical bases), the number of 5 nucleotides incorporated of the complementary base will reflect the number present in the homopolymeric region. Accordingly, determining the number of nucleotides incorporated for each nucleotide addition, will reveal this information, and hence contribute to the pattern of 10 nucleotide incorporation.

Hence, a primer extension protocol may involve annealing a primer as described above, adding a nucleotide, performing a polymerase-catalysed primer extension reaction, detecting the presence or absence of 15 incorporation of said nucleotide (and advantageously also determining the amount of each nucleotide incorporated) and repeating the nucleotide addition and primer extension steps etc. one or more times. As discussed above, single (i.e. individual) nucleotides 20 may be added successively to the same primer-template mixture, or to separate aliquots of primer-template mixture, etc. according to choice.

In order to permit the repeated or successive (iterative) addition of nucleotides in a primer- 25 extension procedure, the previously-added nucleotide must be removed. This may be achieved by washing, or more conveniently, by using a nucleotide-degrading enzyme, for example as described in detail in WO98/28440.

Accordingly, in a principal embodiment of the present invention, a nucleotide degrading enzyme is used to degrade any unincorporated or excess nucleotide. Thus, if a nucleotide is added which is not incorporated (because it is not complementary to the target base), or 35 any added nucleotide remains after an incorporation event (i.e. excess nucleotides) then such unincorporated nucleotides may readily be removed by using a

- 21 -

nucleotide-degrading enzyme. This is described in detail in WO98/28440.

The term "nucleotide degrading enzyme" as used herein includes any enzyme capable of specifically or 5 non-specifically degrading nucleotides, including at least nucleoside triphosphates (NTPs), but optionally also di- and mono-phosphates, and any mixture or combination of such enzymes, provided that a nucleoside triphosphatase or other NTP-degrading activity is 10 present. Where a chain terminating nucleotide is used (e.g. a dideoxy nucleotide is used), the nucleotide degrading enzyme should also degrade such a nucleotide. Although nucleotide-degrading enzymes having a 15 phosphatase activity may conveniently be used according to the invention, any enzyme having any nucleotide or nucleoside degrading activity may be used, e.g. enzymes which cleave nucleotides at positions other than at the phosphate group, for example at the base or sugar residues. Thus, a nucleoside triphosphate degrading 20 enzyme is essential for the invention. Nucleoside di- and/or mono-phosphate degrading enzymes are optional and may be used in combination with a nucleoside triphosphate degrading enzyme.

The preferred nucleotide degrading enzyme is 25 apyrase, which is both a nucleoside diphosphatase and triphosphatase, catalysing the reactions  $\text{NTP} \rightarrow \text{NDP} + \text{Pi}$  and  $\text{NDP} \rightarrow \text{NMP} + \text{Pi}$  (where NTP is a nucleoside triphosphate, NDP is a nucleoside diphosphate, NMP is a nucleotide monophosphate and Pi is inorganic phosphate). 30 Apyrase may be obtained from the Sigma Chemical Company. Other possible nucleotide degrading enzymes include Pig Pancreas nucleoside triphosphate diphosphorydrolase (Le Bel *et al.*, 1980, *J. Biol. Chem.*, 255, 1227-1233). Further enzymes are described in the literature.

35 The nucleotide-degrading enzyme may conveniently be included during the polymerase (i.e. primer extension) reaction step. Thus, for example the polymerase

reaction may conveniently be performed in the presence of a nucleotide-degrading enzyme. Although less preferred, such an enzyme may also be added after nucleotide incorporation (or non-incorporation) has 5 taken place, i.e. after the polymerase reaction step.

Thus, the nucleotide-degrading enzyme (e.g. apyrase) may be added to the polymerase reaction mixture (i.e. target nucleic acid, primer and polymerase) in any convenient way, for example prior to or simultaneously 10 with initiation of the reaction, or after the polymerase reaction has taken place, e.g. prior to adding nucleotides to the sample/primer/polymerase to initiate the reaction, or after the polymerase and nucleotide are added to the sample/primer mixture.

15 Conveniently, the nucleotide-degrading enzyme may simply be included in the reaction mixture for the polymerase reaction, which may be initiated by the addition of the nucleotide.

20 According to the present invention, detection of nucleotide incorporation can be performed in a number of ways, such as by incorporation of labelled nucleotides which may subsequently be detected, or by using labelled probes which are able to bind to the extended sequence.

25 The method may be performed using a sanger sequencing method combined with a standard detection strategy, e.g. electrophoresis or mass spectrometry to analyse, or determine, nucleotide incorporation. However, it is preferred to use a sequencing-by-synthesis method, due to the fact that the extension 30 reactions are quantitative, i.e. that the nucleotide incorporation may be determined quantitatively. As mentioned above, sequencing-by-synthesis methods are disclosed extensively in US-A-4,863,849, which discloses a number of ways in which activated nucleotide incorporation may be determined or detected, e.g. spectrophotometrically or by fluorescent detection 35 techniques, for example by determining the amount of

- 23 -

nucleotide remaining in the added nucleotide feedstock, following the nucleotide incorporation step. In a sequencing-by-synthesis reaction, determination of the pattern of nucleotide incorporation occurs

5 simultaneously with primer extension. One working definition of Sequencing by synthesis is a method in which a single activated (i.e. labelled) nucleotide is or is not incorporated into a primed template, incorporation being detected by any suitable means.

10 This step is repeated by addition of a different activated nucleotide and incorporation is again detected. These steps are repeated and from the sum of incorporated nucleic acids the sequence can be deduced. The preferred method of sequencing-by-synthesis is

15 however a pyrophosphate detection-based method.

Preferably, therefore, nucleotide incorporation is detected by detecting PPi release, preferably by luminometric detection, and especially by bioluminometric detection.

20 PPi can be determined by many different methods and a number of enzymatic methods have been described in the literature (Reeves *et al.*, (1969), Anal. Biochem., 28, 282-287; Guillory *et al.*, (1971), Anal. Biochem., 39, 170-180; Johnson *et al.*, (1968), Anal. Biochem., 15, 273; Cook *et al.*, (1978), Anal. Biochem. 91, 557-565; and Drake *et al.*, (1979), Anal. Biochem. 94, 117-120).

25 It is preferred to use luciferase and luciferin in combination to identify the release of pyrophosphate since the amount of light generated is substantially proportional to the amount of pyrophosphate released which, in turn, is directly proportional to the amount of nucleotide incorporated. The amount of light can readily be estimated by a suitable light sensitive device such as a luminometer. Thus, luminometric methods offer the advantage of being able to be

30 quantitative.

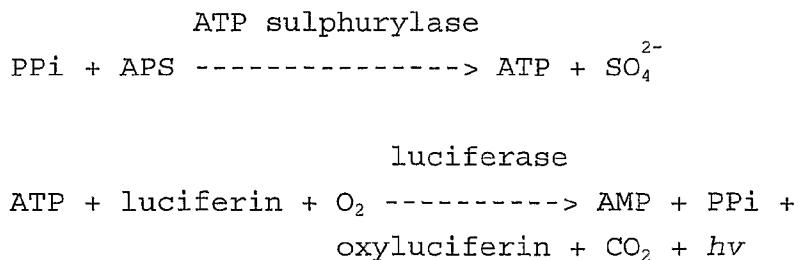
35

Luciferin-luciferase reactions to detect the

- 24 -

release of PPi are well known in the art. In particular, a method for continuous monitoring of PPi release based on the enzymes ATP sulphurylase and luciferase has been developed (Nyrén and Lundin, Anal.

5 Biochem., 151, 504-509, 1985; Nyrén P., Enzymatic method  
for continuous monitoring of DNA polymerase activity  
(1987) Anal. Biochem Vol 167 (235-238)) and termed ELIDA  
(Enzymatic Luminometric Inorganic Pyrophosphate  
Detection Assay). The use of the ELIDA method to detect  
10 PPi is preferred according to the present invention.  
The method may however be modified, for example by the  
use of a more thermostable luciferase (Kaliyama et al.,  
1994, Biosci. Biotech. Biochem., 58, 1170-1171) and/or  
ATP sulfurylase (Onda et al., 1996, Bioscience,  
15 Biotechnology and Biochemistry, 60:10, 1740-42). This  
method is based on the following reactions:



25 (APS = adenosine 5'-phosphosulphate)

Reference may also be made to WO 98/13523 and WO 98/28448, which are directed to pyrophosphate detection-based sequencing procedures, and disclose PPi detection methods which may be of use in the present invention.

In a PPi detection reaction based on the enzymes ATP sulphurylase and luciferase, the signal (corresponding to PPi released) is seen as light. The generation of the light can be observed as a curve known as a Pyrogram™. Light is generated by luciferase action on the product, ATP (produced by a reaction between PPi and APS (see below) mediated by ATP sulphurylase) and,

- 25 -

where a nucleotide-degrading enzyme such as apyrase is used, this light generation is then "turned off" by the action of the nucleotide-degrading enzyme, degrading the ATP which is the substrate for luciferase. The slope of 5 the ascending curve may be seen as indicative of the activities of DNA polymerase (PPi release) and ATP sulphurylase (generating ATP from the PPi, thereby providing a substrate for luciferase). The height of the signal is dependent on the activity of luciferase, 10 and the slope of the descending curve is, as explained above, indicative of the activity of the nucleotide-degrading enzyme. As explained below, Pyrogram™ in the context of a homopolymeric region, peak height is also 15 indicative of the number of nucleotides incorporated for a given nucleotide addition step. Then, when a nucleotide is added, the amount of PPi released will depend upon how many nucleotides (i.e. the amount) are incorporated, and this will be reflected in the slope height.

20 Advantageously, by including the PPi detection enzyme(s) (i.e. the enzyme or enzymes necessary to achieve PPi detection according to the enzymatic detection system selected, which in the case of ELIDA, will be ATP sulphurylase and luciferase) in the 25 polymerase reaction step, the method of the invention may readily be adapted to permit extension reactions to be continuously monitored in real-time, with a signal being generated and detected, as each nucleotide is incorporated.

30 Thus, the PPi detection enzymes (along with any enzyme substrates or other reagents necessary for the PPi detection reaction) may simply be included in the polymerase reaction mixture.

35 A potential problem which has previously been observed with PPi-based sequencing methods is that dATP, used in the chain extension reaction, interferes in the subsequent luciferase-based detection reaction by acting

as a substrate for the luciferase enzyme. This may be reduced or avoided by using, in place of deoxyadenosine triphosphate (ATP), a dATP analogue which is capable of acting as a substrate for a polymerase but incapable of 5 acting as a substrate for a PPi-detection enzyme. Such a modification is described in detail in WO98/13523.

The term "incapable of acting" includes also 10 analogues which are poor substrates for the detection enzymes, or which are substantially incapable of acting as substrates, such that there is substantially no, negligible, or no significant interference in the PPi detection reaction.

Thus, a further preferred feature of the invention 15 is the use of a dATP analogue which does not interfere in the enzymatic PPi detection reaction but which nonetheless may be normally incorporated into a growing DNA chain by a polymerase. By "normally incorporated" is meant that the nucleotide is incorporated with normal, proper base pairing. In the preferred 20 embodiment of the invention where luciferase is a PPi detection enzyme, the preferred analogue for use according to the invention is the [1-thio]triphosphate (or  $\alpha$ -thiotriphosphate) analogue of deoxy ATP, preferably deoxyadenosine [1-thio]triphosphate, or 25 deoxyadenosine  $\alpha$ -thiotriphosphate (dATP $\alpha$ S) as it is also known. dATP $\alpha$ S, along with the  $\alpha$ -thio analogues of dCTP, dGTP and dTTP, may be purchased from Amersham Pharmacia. Experiments have shown that substituting dATP with 30 dATP $\alpha$ S allows efficient incorporation by the polymerase with a low background signal due to the absence of an interaction between dATP $\alpha$ S and luciferase. False signals are decreased by using a nucleotide analogue in place of dATP, because the background caused by the 35 ability of dATP to function as a substrate for luciferase is eliminated. In particular, an efficient incorporation with the polymerase may be achieved while the background signal due to the generation of light by

the luciferin-luciferase system resulting from dATP interference is substantially decreased. The dNTP $\alpha$ S analogues of the other nucleotides may also be used in place of the other dNTPs.

5 Another potential problem which has previously been observed with sequencing-by-synthesis methods is that false signals may be generated and homopolymeric stretches (i.e. CCC) are difficult to sequence with accuracy. This may be overcome by the addition of a  
10 single-stranded nucleic acid binding protein (SSB) once the extension primers have been annealed to the template nucleic acid. The use of SSB in sequencing-by-synthesis is discussed in WO 00/43540 of Pyrosequencing AB.

It will be understood that in the method of the  
15 invention, differing amounts of nucleic acid template may be present when multiple nucleic acid molecules are to be typed. In order to be able to quantify the number of nucleotides incorporated upon addition in certain embodiments, it is preferred to design the primers and  
20 nucleotide dispensation in such a way that a reference signal is generated for each primer which corresponds to a single nucleotide incorporation event. The reference signal is generated in the absence of nucleotide incorporation in the other primer-extension reactions.  
25 The reference signal allows for calibration of the signals relating to the same template. The reference peaks are clearly shown on Figure 9, and the height of the variant site signal can be correlated to the reference signal to increase accuracy.

30 The step of detecting nucleotide incorporation by detecting PPi release results in a signal indicative of the amount of pyrophosphate released, and hence the amount of nucleotide incorporated. In the method of the invention, 2 or more distinct primers are used  
35 sequentially or simultaneously in a primer-extension reaction. Thus, in the case of the simultaneously added primers, for every nucleotide addition, 0, 1 or more

nucleotides may be incorporated into the growing DNA chains. The signal generated in the pyrophosphate detection step will therefore be indicative of the number of nucleotides incorporated in the primer-5 extension step for the combination of all primers bound to the template DNA. The size of the signal (i.e. the height of each peak) can therefore be correlated directly to the number of incorporated nucleotides. In certain embodiments, the primer needs only to be 10 subjected to 1 to 20, preferably 1 to 10, e.g. 1 to 5 and most preferably 1 to 4 cycles of nucleotide addition.

In one embodiment of the invention, 2 or more primers are hybridized (simultaneously) at, adjacent or 15 near to variable sites in the target nucleic acid. Each primer being responsible for the typing of one or possibly more variable sites. Primer extension is then performed as described above, and primer extension occurs for each primer only if the nucleotide added is 20 complementary to the target base. Thus, when 2 primers are used simultaneously, none, 1, 2 or more (for homopolymeric regions) nucleotide incorporation events may occur upon the addition of any given nucleotide. The primer extension reaction is carried out simultaneously 25 for all hybridized primers in the reaction mixture. Thus, the detected nucleotide incorporation gives a cumulative picture for all hybridized primers. In this manner, the pattern of nucleotide incorporation may be directly determined. Preferably, when an 30 extension reaction extends across a variable site, nucleotide incorporation occurs only at that site.

In a further embodiment of the invention, the primers may be added sequentially to the primer extension reaction.

35 In this case, the pattern of nucleotide incorporation may be determined for each primer separately, and then "added together" to obtain a

5 cumulative picture/pattern. In a modified version of this embodiment of the invention, the first primer is hybridized to the target nucleic acid, undergoes a primer extension reaction, which is terminated after the variable site has been sequenced, by the addition of a chain terminator. Chain terminators are well known in the art, and include dideoxynucleotides. A second primer is then added to sequence a second variable site, and the sequencing is again terminated by the addition 10 of a chain terminator. This method may be repeated until all variable regions of interest have been sequenced.

15 In a further particularly preferred embodiment which is also discussed above and in the Examples, the extension primers are hybridized to the template, and the primers are extended simultaneously. The primers are designed to enable primer extension to occur over the variable sites sequentially - i.e. primer extension occurs for each primer simultaneously, but primer 20 extension over a variable site occurs in turn, whilst the other primers are extended over a conserved/semiconserved region or more preferably are not extended at all due to the addition of non-complementary nucleotides. The pattern of nucleotide addition is 25 preferably pre-determined to allow extension of the primers to occur sequentially over the variable sites. The primers may bind 1 to 40, 1 to 20, 1 to 10, 1 to 5 nucleotides from or adjacent to the variable site.

30 Optionally once a primer has been extended over a variable site, a chain terminator, such as a dideoxynucleotide, may be added to specifically terminate the chain extension reaction of that primer. It will be understood that nucleotide incorporation signals will be generated for all primers during the 35 primer extension reaction, and will contribute to the pattern obtained. Nevertheless, different regions of the pattern will preferably relate to just one of the

variable sites.

In a still further modified embodiment of the invention, chain terminators may be employed in place of dNTPs or in combination with dNTPs, using simultaneously 5 hybridised primers. In this case, the primers are selected or designed to ensure that primer extension from each primer takes place sequentially, i.e. that nucleotides are first incorporated from the first primers, the first extension reaction is complete, 10 before nucleotide incorporation from the next primer takes place. This embodiment also requires that the nucleotides are added in predetermined order.

Indeed, so-called "intelligent" primer design may be used to carry out the method of the invention in a 15 desired or pre-selected (i.e. predetermined) manner. This may be applied both to the number of extension primers employed, and to the design of the sequence thereof. "Intelligent" primer design is optimally performed with an "intelligent" order of addition of 20 nucleotides to enable the sequencing of the individual variable sites to be performed in isolation. Such 'intelligent' design of primers and the order of nucleotide addition is described in more detail in the Examples.

25 The method of the invention may conveniently be performed in a single reaction vessel, whether a "simultaneous" or "sequential" primer extension embodiment is used. Thus, for example, all extension primers may be added together, or sequentially into a 30 single reaction vessel.

In order for the primer-extension reaction to be performed, the nucleic acid molecule, regardless of whether or not it has been amplified, is conveniently provided in a single-stranded format. The nucleic acid 35 may be subjected to strand separation by any suitable technique known in the art (e.g. Sambrook *et al.*, *supra*), for example by heating the nucleic acid, or by

- 31 -

heating in the presence of a chemical denaturant such as formamide, urea or formaldehyde, or by use of alkali.

However, this is not absolutely necessary and a double-stranded nucleic acid molecule may be used as 5 template, e.g. with a suitable polymerase having strand displacement activity.

Where a preliminary amplification step is used, regardless of how the nucleic acid has been amplified, all components of the amplification reaction need to be 10 removed, to obtain pure nucleic acid, prior to carrying out the typing assay of the invention. For example, unincorporated nucleotides, PCR primers, and salt from a PCR reaction need to be removed. Methods for purifying nucleic acids are well known in the art (Sambrook et al., 15 supra), however a preferred method is to immobilize the nucleic acid molecule, removing the impurities via washing and/or sedimentation techniques.

Optionally, therefore, the target nucleic acid may be provided with a means for immobilization, which may be introduced during amplification, either through the 20 nucleotide bases or the primer/s used to produce the amplified nucleic acid.

To facilitate immobilization, the amplification primers used according to the invention may carry a 25 means for immobilization either directly or indirectly. Thus, for example the primers may carry sequences which are complementary to sequences which can be attached directly or indirectly to an immobilizing support or may carry a moiety suitable for direct or indirect 30 attachment to an immobilizing support through a binding partner.

Numerous suitable supports for immobilization of 35 DNA and methods of attaching nucleotides to them, are well known in the art and widely described in the literature. Thus for example, supports in the form of microtitre wells, tubes, dipsticks, particles, fibres or capillaries may be used, made for example of agarose,

cellulose, alginate, teflon, latex or polystyrene. Advantageously, the support may comprise magnetic particles e.g. the superparamagnetic beads produced by Dynal AS (Oslo, Norway) and sold under the trademark 5 DYNABEADS. Chips may be used as solid supports to provide miniature experimental systems as described for example in Nilsson et al. (Anal. Biochem. (1995), 224:400-408).

The solid support may carry functional groups such 10 as hydroxyl, carboxyl, aldehyde or amino groups for the attachment of the primer or capture oligonucleotide. These may in general be provided by treating the support to provide a surface coating of a polymer carrying one 15 of such functional groups, e.g. polyurethane together with a polyglycol to provide hydroxyl groups, or a cellulose derivative to provide hydroxyl groups, a polymer or copolymer of acrylic acid or methacrylic acid to provide carboxyl groups or an amino alkylated polymer 20 to provide amino groups. US patent No. 4,654,267 describes the introduction of many such surface coatings.

Alternatively, the support may carry other moieties 25 for attachment, such as avidin or streptavidin (binding to biotin on the nucleotide sequence), DNA binding proteins (e.g. the lac I repressor protein binding to a lac operator sequence which may be present in the primer or oligonucleotide), or antibodies or antibody fragments 30 (binding to haptens e.g. digoxigenin on the nucleotide sequence). The streptavidin/biotin binding system is very commonly used in molecular biology, due to the relative ease with which biotin can be incorporated 35 within nucleotide sequences, and indeed the commercial availability of biotin-labelled nucleotides. This represents one preferred method for immobilisation of target nucleic acid molecules according to the present invention. Streptavidin-coated DYNABEADS are commercially available from Dynal AS.

As mentioned above, immobilization may conveniently take place after amplification. To facilitate post amplification immobilisation, one or both of the amplification primers are provided with means for 5 immobilization. Such means may comprise as discussed above, one of a pair of binding partners, which binds to the corresponding binding partner carried on the support. Suitable means for immobilization thus include biotin, haptens, or DNA sequences (such as the lac 10 operator) binding to DNA binding proteins.

When immobilization of the amplification products is not performed, the products of the amplification reaction may simply be separated by for example, taking them up in a formamide solution (denaturing solution) 15 and separating the products, for example by electrophoresis or by analysis using chip technology. Immobilization provides a ready and simple way to generate a single-stranded template for the extension reaction. As an alternative to immobilization, other 20 methods may be used, for example asymmetric PCR, exonuclease protocols or quick denaturation/annealing protocols on double stranded templates may be used to generate single stranded DNA. Such techniques are well known in the art.

25 The method of the invention allows the typing (e.g. genotyping) of one or more nucleic acid molecule derived from an individual (e.g. a patient under clinical test, a tissue sample for typing, or a microorganism for identification). Thus, the method of the invention is 30 capable of distinguishing between different genotypes within a species. This is particularly useful in the field of identification of microbial species, where many genotypes of one microbe may exist, for example, there are currently seven known genotypes of the Hepatitis C 35 Virus.

The method of the present invention is particularly advantageous in the diagnosis of pathological conditions

characterised by the presence of specific DNA, particularly latent infectious diseases such as viral infection e.g. by herpes, hepatitis or HIV. Also, the method can be used to characterise or type and quantify 5 bacterial, protozoal and fungal infections where samples of an injecting organism may be difficult to obtain or where an isolated organism is difficult to grow *in vitro* for subsequent characterisation as in the case of *P. falciparum* or *Chlamydia* species. Due to the simplicity 10 and speed of the method it may also be used to detect other pathological agents which cause diseases such as syphilis and meningitis. Even in cases where samples of the injecting organism may be easily obtained, the speed of this method compared with overnight incubation of a 15 culture may make the method according to the invention preferable over conventional techniques.

The method of the present invention may be used to analyse two or more single nucleotide polymorphisms (SNPs) within one or more genes, or two or more genes, 20 in an individual. Many diseases and conditions may be associated with (or linked to) combinatorial polymorphisms within the same gene, or within distinct genes. For example, in WO 00/22166, it has been suggested that a combination of SNPs within several 25 genes gives a polymorphic pattern which may be used to predict the likelihood of cardiovascular disease, allowing detailed prognosis for an individual, and predicting whether a particular therapeutic regime would be effective in improving a cardiovascular condition.

30 Thus, the method of the invention can be used to give a quick prognosis on the particular genotype of an individual, allowing tailored therapy to be administered. Example 2 shows that multiplex genotyping can be performed for SNPs in the RAAS system. In this 35 example, one nucleic acid contains 2 SNPs (EU7) and two additional nucleic acids contain 1 SNP each (EU8 and EU11).

- 35 -

The method of the invention is advantageous in that it determines the exact sequence of the variable sites (i.e. is based on a sequencing procedure, it avoids costly and cumbersome procedures, such as 5 electrophoresis, and advantageously labelled nucleotides and/or primers, and large numbers of samples can be analysed in a short time.

The primer extension reaction generates a "pattern" or "fingerprint" indicative of nucleotide incorporation, 10 correlated to the nucleotide added to the reaction mixture. The pattern is a cumulative picture of nucleotide incorporation for the primers designed to detect nucleotide incorporation at 2 or more variable sites within the target nucleic acid molecule(s). To 15 enable the target nucleic acid molecule(s) to be typed, reference patterns are used, using the same variable sites and extension primers. Each genotype should produce a different pattern, facilitating identification by comparison to the reference pattern which can be 20 determined theoretically.

The method of the invention relies upon the knowledge of the location and nature of the variable sites, together with further known sequence information (e.g. with known sequences of conserved/semi-conserved 25 regions) from which to determine an appropriate primer binding site and design a complementary extension primer. Using the method of the invention, any combination of variable sites may be used in the typing method. It will be understood by those skilled in the 30 art that the method of the invention is not limited to multiple variable sites within genes, but the method is also applicable to non-coding regions. The pattern may be obtained for variable sites which are in one or more of the same gene, in related genes, in disparate genes, 35 or in non-coding regions.

The invention also comprises kits for carrying out the method of the invention. These will normally

include one or more of the following components:

optionally primer(s) for *in vitro* amplification; two or more primers for the primer extension reaction; nucleotides for amplification and/or for the primer extension reaction (as described above); a polymerase enzyme for the amplification and/or primer extension reaction; and means for detecting primer extension (e.g. means of detecting the release of pyrophosphate as outlined and defined above).

10 In certain embodiments, the kit will also include instructions for the order of addition of the nucleotides.

15 The invention will now be described by way of non-limiting examples with reference to the drawings in which:-

Figure 1 shows schematically one method for the typing of nucleic acid using multiple primers (multiplexing) simultaneously in a primer extension reaction;

20 Figure 2 shows the sequence of the 5' untranslated region (5'-UTR) of seven Hepatitis C virus (HCV) genotypes, wherein the arrows indicate the positions of the amplification and extension primers, and the nucleotides highlighted in bold type illustrate the 25 variable region to be sequenced by the primer extension reaction;

30 Figure 3 shows theoretical traces which would be obtained (light generated (indicating nucleotide incorporation) versus time (and nucleotide addition)), for the seven genotypes of HCV studied. The experimental conditions and extension primers theoretically used are described in Example 1. Three distinct extension primers were theoretically used simultaneously in a primer-extension reaction mixture.

35 Inorganic pyrophosphate PPi is released in a DNA-polymerase catalyzed reaction if a nucleotide is incorporated. The PPi is monitored by coupled enzymatic

- 37 -

reactions using ATP sulphurylase and luciferase. Light generated as a result is measured by a CCD detector or luminometer;

5 Figure 4 shows traces (light generated (indicating nucleotide incorporation) versus time (and nucleotide addition)), obtained for six samples containing different HCV genotypes. The experimental conditions and primers used are described in Example 1. The incorporation of a nucleotide into the extending primer 10 results in the release of PPi, which is detected using a coupled enzymatic reactions using ATP sulphurylase and luciferase. Light generated as a result of successful extension is measured by a CCD camera or luminometer;

15 Figure 4a shows the trace obtained for HCV genotype 1a;

Figure 4b shows the trace obtained for HCV genotype 1b;

Figure 4c shows the trace obtained for HCV genotype 2a;

20 Figure 4d shows the trace obtained for HCV genotype 2b;

Figure 4e shows the trace obtained for HCV genotype 3a;

25 Figure 4f shows the trace obtained for HCV genotype 3b;

Figure 5 shows three potential primer binding positions for the SNP Eu6 from the ACE gene (Angiotensin Converting Enzyme). Figure 5a shows a primer (boxed nucleotides) bound to the template with its 3' end 4 30 nucleotides from the SNP position, Figure 5b shows a primer bound to the template with the 3' and 5 nucleotides from the SNP position and in Figure 5c the primer is bound 10 nucleotides from the SNP position. In all figures, the 2 potential variants at the SNP site 35 are shown (G/A in template strand);

Figure 6 shows the theoretical output from Pyrosequencing™ reactions for two SNP positions. The

theoretical output is plotted as nucleotide dispensed into the reaction versus peak height (correlated to light emitted from the Pyrosequencing™ reaction).

5 Figure 6a shows the theoretical output for sequencing G/ACAG, in this case, the primer would be adjacent to the polymorphic position. The theoretical output shown is for the heterozygote (i.e. the individual has one copy of the SNP A and one copy of the SNP G). Figure 6b shows the theoretical output for TGAAC/TA. The primer 10 is thus bound 4 nucleotides away from the SNP. Again, the pattern shown would result from a heterozygous individual (C/T). Figure 6c shows a cumulation of the two individual sequencing reactions in one primer extension reaction mixture;

15 Figure 7 shows a simplified multiplexing analysis wherein the extension primers are designed in such a way that their 3' ends are positioned at different distances from the polymorphic position. This enables the design of an "intelligent" order of addition of nucleotides to 20 be determined to enable the SNP (marked X) to be sequenced in isolation. Thus, the extension primers should be designed in parallel with the dispensation order;

25 Figure 8 shows the theoretical output for five SNPs present in the RAAS system - Eu4 (ACE G2215A), Eu8 (ATG C521T), Eu10 (ATP T573C), Eu6 (ACE T3409C) and Eu3 (ACE T1237C). The theoretical outputs are plotted as in Figure 7. The extension primers are positioned such that the sequence that is analysed for Eu4 is G/A 30 CTGCCTG, Eu8 is CACCA/G TGG, Eu10 is C/TCCGATAGGGC, Eu6 is ACTTC/TG and Eu3 is AGACA/GGGC;

35 Figures 8a and 8b show the theoretical output expected to be obtained when the SNPs Eu4 and Eu8 are typed in a standard one-primer only reaction. The SNP nucleotide incorporation position are framed;

Figure 8a shows the theoretical output when the individual is heterozygous (A/G) and Figure 8b shows the

output expected when the individual is homozygous (G/G). Figure 8c shows the output expected when the two SNPs are sequenced simultaneously in the same reaction (multiplexed). The polymorphic positions are framed;

5 Figures 8d, 8e and 8f are the theoretical results obtained from sequencing Eu10, Eu6 and Eu3 alone, respectively. SNP Eu10 and Eu6 are shown as heterozygotes (C/T and C/T, respectively) and Eu3 as a homozygote (A/A). The theoretical patterns for the 3  
10 SNPs are combined in Figure 8g, and the SNP positions are framed;

15 Figure 9 shows the results obtained as traces (light generated (indicating nucleotide incorporation) versus time and nucleotide incorporation for seven reactions containing differing templates and primer combinations. The experimental conditions and primers used are described in Example 2. The incorporation of a nucleotide into the extending primer results in the release of PPi, which is detected using a coupled  
20 enzymatic reactions using ATP sulphurylase and luciferase. Light generated as a result of successful extension is measured by a CCD camera or luminometer;

25 Figure 9a shows the trace obtained for Eu4;

Figure 9b shows the trace obtained for Eu8;

30 Figure 9c shows the trace obtained for Eu4 and Eu8 simultaneously sequenced;

Figure 9d shows the trace obtained for Eu10;

Figure 9e shows the trace obtained for Eu6;

Figure 9f shows the trace obtained for Eu3;

35 Figure 9g shows the trace obtained for Eu10, Eu6 and Eu3, simultaneously sequenced;

Figure 10 shows the theoretical output for SNPs Eu8, Eu7 and Eu11 present in the RAAS system. The theoretical outputs are plotted as in Figure 7. The sequences analyzed are Eu8 CACCA/GTGGACAG, Eu7 T/CGGCCGGGTACGAG/TG and Eu11 GAGCA/GTTAG. Therefore, the fragment Eu7 contains two polymorphic sites;

- 40 -

Figure 10a shows the theoretical trace for Eu8 (G/G) ;

Figure 10b shows the theoretical trace for Eu7 (C/C and T/T) ;

5 Figure 10c shows the theoretical trace for Eu11 (A/G) ; and

Figure 10d shows the multiplex theoretical trace for Eu7, Eu8 and Eu11;

10 Figure 11 shows the result obtained as a trace (light generated versus nucleotide addition) for the multiplex reaction as defined in Example 3. The polymorphic positions are framed and the reference peaks shown (arrows). The genotype for the individual typed is Eu8 G/G, Eu7 C/C and T/T Eu11 A/G;

15 Figure 12 is a scheme showing primer design for 3 separate nucleic acid fragments, for the Plasminogen Activator Inhibitor I gene, the Prothrombin gene and the Factor V gene. The arrows marked with a \* correspond to the sequencing primers and the outer arrows correspond 20 to the PCR primers. The biotinylated primer is indicated with a B. The polymorphic position in the gene of interest is marked by an X;

25 Figure 13 shows the expected output from Pyrosequencing™ reactions for SNPs in Plasminogen-activator inhibitor 1 (4G/5G deletion), Prothrombin (G20210A) and Factor V (G1691A). The theoretical output is shown as for Figure 7. The sequence analysed for PAI1 is (C)ACGTG, Prothrombin is GCTC/TGCTGA and Factor V AGGCA/GAGGAA;

30 Figure 13a shows the theoretical trace for PAI1 (C/C - no deletion) ;

Figure 13b shows the theoretical trace for Prothrombin (C/T) ;

35 Figure 13c shows the theoretical trace for Factor V (A/G) ;

Figure 13d shows the theoretical trace for a combination of the three SNPs in one multiplex reaction;

- 41 -

Figure 13e shows the theoretical trace for the genotype PAI1 C/C (no deletion), Prothrombin C/C and Factor V G/G;

5 Figure 13f shows the theoretical trace for the genotype PAI1 double deletion, Prothrombin C/C and Factor V A/A;

Figure 13g shows the theoretical trace for the genotype PAI1 del/C, Prothrombin T/T and Factor V G/G;

10 Figure 13h shows the theoretical trace for the genotype PAI1 C/C (no deletion), Prothrombin T/C and Factor V G/G;

Figure 13i shows the theoretical trace for the genotype PAI1 C/del, Prothrombin C/C and Factor V G/G; and

15 Figure 13j shows the theoretical trace for the genotype PAI1 C/C (no deletion), Prothrombin C/C and Factor V A/G;

20 Figure 14 shows the results obtained as traces (light generated versus nucleotide addition) for six reactions. The materials and methods are described in Example 4;

25 Figure 14a shows the trace obtained for the genotype PAI1 C/C (no deletion), Prothrombin C/C and Factor V G/G and corresponds to the theoretical pattern shown on Figure 13e;

Figure 14b shows the trace obtained for the genotype PAI1 double deletion, Prothrombin C/C and Factor V A/A and corresponds to the theoretical pattern shown on Figure 13f;

30 Figure 14c shows the trace obtained for the genotype PAI1 del/C, Prothrombin T/T and Factor V G/G and corresponds to the theoretical pattern shown on Figure 13g;

35 Figure 14d shows the trace obtained for the genotype PAI1 C/C (no deletion), Prothrombin T/C and Factor V G/G and corresponds to the theoretical pattern shown on Figure 13h;

- 42 -

Figure 14e shows the trace obtained for the genotype PAI1 C/del, Prothrombin C/C and Factor V G/G and corresponds to the theoretical pattern shown on Figure 13i; and

5 Figure 14f shows the trace obtained for the genotype PAI1 C/C (no deletion), Prothrombin C/C and Factor V A/G and corresponds to the theoretical pattern shown on Figure 13j;

10 Figure 15 depicts the localisation of the primers with regard to the CYP2D6 gene. A segment of the gene with particular highlighted polymorphisms can be seen at the top of this figure. The 61118 and 2162 fragments, as amplified by nested PCR primers are at the bottom of the figure. The extension primers used for the 15 multiplexing reactions are shown above the 61118 and 2162 fragments;

Figure 16 represents the theoretical output obtained for two genotypes of the CYP2D6 gene. The traces were calculated as described previously;

20 Figure 16a shows the theoretical output for G1934 A (A/G), G 1749 C (C/G), T1795 del (no deletion) and G 1846 T (T/g);

25 Figure 16b shows the theoretical output for G1934 A (A/G), G 1749 C (C/G), T1795 del T/deletion and G1846T (T/G); and

30 Figure 17 shows the result obtained from Example 5 as a trace (light generated versus nucleotide added). The experimental conditions are described in Example 5. The genotype of the individual typed is G1934A G/G, G1749C G/G, T1795 del T/T (no deletion) G1846T G/G. Also shown is the theoretical output plot for this genotype.

#### EXAMPLE 1

35

#### Serum samples

72 sera from HCV-positive Veterans were obtained from Stanford Veteran hospital. 10 HCV-positive sera were obtained from Iran.

5       Synthesis and purification of oligonucleotides

The oligonucleotides HCV-PCR-OUTF  
(5'- CCCTGTGAGGAACTWCTGTCTTCACGC), HCV-PCR-OUTR  
(5'-GCTCATGRTGCACGGTCTACGAGACCT), HCV-PCR-INF  
10      (5'-TCTAGCCATGGCGTTAGTAYGAGTGT), BHCV-PCR-INR  
(5'-Biotin-CACTCGCAAGCACCCTATCAGGCAGT), HCV-SEQF1  
(5'-GGAACCGGTGAGTACACCGGAAT), HCV-SEQF2  
(5'-GACYGGGTCCCTTCTTGG), HCV-SEQF3  
(5'-ATTGGGGCGTGCCCCCGC), were all synthesized and HPLC  
15      purified by MWG Biotech (High points, NC, USA).

RNA extraction, cDNA synthesis and amplification

RNA was extracted from 100  $\mu$ l of patient sera using  
20      Ambion's Totally RNA isolation kit ([www.ambion.com](http://www.ambion.com),  
Ambion (Europe) Ltd., Cambridge, UK). cDNA was  
synthesized using the kit Superscript<sup>TM</sup> Preamplification  
system from Invitrogen ([www.invitrogen.com](http://www.invitrogen.com), Invitrogen  
Ltd., Paisley, UK). First strand cDNA synthesis  
25      employed an RNA/primer mixture containing, 5  $\mu$ l RNA and  
1  $\mu$ l 0.5  $\mu$ g/ $\mu$ l Oligo (dT) random primer which was  
incubated at 70°C for 10 min and then placed on ice for  
at least 1 min. A reaction mixture containing 2  $\mu$ l 10 X  
PCR buffer (200mM Tris-HCl (pH 8.4), 500 mM KCl), 2  $\mu$ l  
30      25 mM MgCl<sub>2</sub> 10 mM dNTP mix and 0.1 M DTT, was added to  
each RNA/primer mixture, mixed gently collected by brief  
centrifugation and then incubated at 42°C for 5 min.  
Two hundred units of Superscript II Reverse  
Transcriptase was added to each tube, and incubated at  
35      40°C for 50 min. The reaction was terminated by  
incubating at 70°C for 15 minutes and then chilled on  
ice. The nucleic acid was collected by brief

- 44 -

centrifugation. 1  $\mu$ l of RNase H was added to each tube and incubated for 20 mm at 37°C. Outer PCR was performed on 1  $\mu$ l of cDNA using HCV-PCR-OUTF and HCV-PCR-OUTR PCR. The outer PCR was diluted by 500,000 5 times and 1  $\mu$ l of that was used as a template for inner PCR using primers HCV-PCR-INF and HCV-PCR-INR.

Template Preparation

The biotinylated PCR products were immobilized onto streptavidin-coated super paramagnetic beads Dynabeads™ 5 M280-Streptavidin (Dynal Biotech ASA, Oslo, Norway). Single-stranded DNA was obtained by discarding the supernatant after incubation of the immobilized PCR product in 0.10 M NaOH for 3 min. Five pmol of sequencing primers HCV-SEQF1, HCV-SEQF2, and HCV-SEQF3 10 were hybridized to the immobilized strand, as described in Ronaghi *et al.*, 1996, Analytical Biochemistry, 242, 84-89.

Primer extension reaction

15 The primed DNA templates were placed in a microtiter plate containing 0.5 µg SSB (Amersham Pharmacia Biotech, USA), and Pyrosequencing™ substrates and enzymes (www.pyrosequencing.com Pyrosequencing AB, Uppsala, Sweden) nucleotides were dispensed using fully automated microtiter plate-based PSQ™ Pyrosequencing™ instrument. The sequencing procedure was carried out by stepwise elongation of the primer-strand upon pre-specified addition of four different nucleotides. The template 20 was hybridized with the three extension primers described above. The progress of sequencing was followed in real-time using Pyrosequencing™ Tag software, (Pyrosequencing™ AB, Uppsala, Sweden) and subtyping was performed manually.

25 30 HCV positive blood sera from 89 different patients was collected and HCV RNA was extracted as described above. Subsequent to cDNA synthesis, PCR was performed to amplify a 236-base long region from 5' UR. One of the primers in the PCR was biotinylated. After capture of the PCR products on magnetic beads and template preparation, sequencing-by-synthesis was performed.

Results**Principle of the HCV typing method.**

5 The principle of the typing method described above is outlined in Figure 1. In this model system, extension primers are hybridized to the target sample DNA, which is immobilized on magnetic beads.

10 The extension primers hybridise specifically to the conserved region adjacent to the variable region. In this set of experiments, 3 sequencing primers for HCV were used. The primers and their alignment to the HCV genomes are shown in Figure 2.

15 The signals resulting from the specific extension of each primer are directly correlated to the number of nucleotides incorporated. The 'fingerprint' produced can therefore be used to identify the genotype of the 20 individual, against reference fingerprints, which can be theoretically deduced from the sequences of the variable regions. Reference fingerprints calculated theoretically from the sequence of the variable regions are shown on Figure 3. These can be used to type the 25 results shown on Figure 4: Figure 4a is the fingerprint for HCV 1a, Figure 4b is the fingerprint for HCV 1b, Figure 4c is the fingerprint for HCV 2a, Figure 4d is the fingerprint for HCV 2b, Figure 4e is the fingerprint for HCV 3a, and Figure 4f is the fingerprint for HCV 3b.

30 Therefore, using the method of the invention, it was possible to genotype HCV infection. Of the 77 sera analyzed by the method of the invention. 35% were infected with HCV 1a, 29% with HCV 1b, 21% with HCV 2a, 4% with HCV 2b, 1% with HCV 3a and 10% with HCV 3b. Of 35 the 10 analysed samples from Iran, the following results were obtained; 1a, 1; 1b, 3; 2a, 3; 3a, 2 and 3b; 1.

- 47 -

EXAMPLE 2

Typing of SNPs in the RAAS System

5      Templates and Primers

Genomic DNA was isolated according to standard methods,  
 PCR temples was generated with specific primers  
 according to the table below.

10

PCR fragment	PCR primer 1 5'-biotin	PCR primer 2	From ref.	
Eu3	GGA CCA GCT CTC CAC AGT GC	GCC AGC ACG TCC CCA AT	US 6197505 ACEe8R (PCR2)	
Eu4	GAT TCC CCT CTC CCT GTA CCT	GCC AGG AAG TTT GAT GTG AAC	ACEe15R (PCR1)	
15	Eu6	CTC GCT CTG CTC CAG GTA C	GCC TCC TTG GAC TGG TAG AT	ACEe24F (PCR2)
	Eu8	CCA GGG CAG GGC TGA TA	CAA ACG GCT GCT TCA GGT	ANGe2f3F
	Eu10	CAT TTC TTG GTT TGT TCT TCT GA	GTT TGT GCT TTC CAT TAT GAG TC	AT1e5f3F

20      The following sequencing primers were used in the  
 multiplex reactions:

Eu3	Eu3s	5'- CCC CGA CGC AGG GAG AC-3'
	A062RS	5'-CCC CGA CGC AGG GAG-3'
	A0943S	5'- CCC CGA CGC AGG G -3'
25	Eu4	5'- GAC CTA GAA CGG GCA GC-3'
	A097FS	5'- GTT CAG GAC CTA GAA -3'
	Eu6	5'- CCT CGC TCC GCT CCA GGT A-3'
	A091FS	5'-CTC GCT CTG CTC-3'
	A063FS	5'-CTC GCT CTG CTC CAG GT-3'
30	Eu8	5'- GCT GTG AAC ACG CCC AC-3'
	A060FS	5'-GCT GCT GCT GCT CA-3'
	Eu10 A088FS (Eu10s)	5'- AGA TCC CAA AAT TCA ACC CT-3'

PCR Amplification

35

The target nucleic acid molecules were amplified by PCR,

- 48 -

either by standard PCR or by multiplex PCR.

5 Simplex PCR: A 50  $\mu$ l PCR reaction was set up for each SNP-specific fragment and sample. All fragments were amplified with the AmpliTaq Gold kit (PE Biosystems) and 1.5 mM MgCl<sub>2</sub> according to the following protocol. (Table 1).

Table 1.

10

PCRmix	1x	100x
10xPCRbuffer	5	500
MgCl <sub>2</sub> (25 mM)	3	300
dNTP (2.5 mM)	2.5	250
DMSO	0	0
Primer a (10 $\mu$ M)	1	100
Primer b (10 $\mu$ M)	1	100
TaqGold (5 units/ $\mu$ l)	0.3	30
H <sub>2</sub> O	32.2	3220
Sum:	45	4500

15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995 1000 1005 1010 1015 1020 1025 1030 1035 1040 1045 1050 1055 1060 1065 1070 1075 1080 1085 1090 1095 1100 1105 1110 1115 1120 1125 1130 1135 1140 1145 1150 1155 1160 1165 1170 1175 1180 1185 1190 1195 1200 1205 1210 1215 1220 1225 1230 1235 1240 1245 1250 1255 1260 1265 1270 1275 1280 1285 1290 1295 1300 1305 1310 1315 1320 1325 1330 1335 1340 1345 1350 1355 1360 1365 1370 1375 1380 1385 1390 1395 1400 1405 1410 1415 1420 1425 1430 1435 1440 1445 1450 1455 1460 1465 1470 1475 1480 1485 1490 1495 1500 1505 1510 1515 1520 1525 1530 1535 1540 1545 1550 1555 1560 1565 1570 1575 1580 1585 1590 1595 1600 1605 1610 1615 1620 1625 1630 1635 1640 1645 1650 1655 1660 1665 1670 1675 1680 1685 1690 1695 1700 1705 1710 1715 1720 1725 1730 1735 1740 1745 1750 1755 1760 1765 1770 1775 1780 1785 1790 1795 1800 1805 1810 1815 1820 1825 1830 1835 1840 1845 1850 1855 1860 1865 1870 1875 1880 1885 1890 1895 1900 1905 1910 1915 1920 1925 1930 1935 1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020 2025 2030 2035 2040 2045 2050 2055 2060 2065 2070 2075 2080 2085 2090 2095 2100 2105 2110 2115 2120 2125 2130 2135 2140 2145 2150 2155 2160 2165 2170 2175 2180 2185 2190 2195 2200 2205 2210 2215 2220 2225 2230 2235 2240 2245 2250 2255 2260 2265 2270 2275 2280 2285 2290 2295 2300 2305 2310 2315 2320 2325 2330 2335 2340 2345 2350 2355 2360 2365 2370 2375 2380 2385 2390 2395 2400 2405 2410 2415 2420 2425 2430 2435 2440 2445 2450 2455 2460 2465 2470 2475 2480 2485 2490 2495 2500 2505 2510 2515 2520 2525 2530 2535 2540 2545 2550 2555 2560 2565 2570 2575 2580 2585 2590 2595 2600 2605 2610 2615 2620 2625 2630 2635 2640 2645 2650 2655 2660 2665 2670 2675 2680 2685 2690 2695 2700 2705 2710 2715 2720 2725 2730 2735 2740 2745 2750 2755 2760 2765 2770 2775 2780 2785 2790 2795 2800 2805 2810 2815 2820 2825 2830 2835 2840 2845 2850 2855 2860 2865 2870 2875 2880 2885 2890 2895 2900 2905 2910 2915 2920 2925 2930 2935 2940 2945 2950 2955 2960 2965 2970 2975 2980 2985 2990 2995 3000 3005 3010 3015 3020 3025 3030 3035 3040 3045 3050 3055 3060 3065 3070 3075 3080 3085 3090 3095 3100 3105 3110 3115 3120 3125 3130 3135 3140 3145 3150 3155 3160 3165 3170 3175 3180 3185 3190 3195 3200 3205 3210 3215 3220 3225 3230 3235 3240 3245 3250 3255 3260 3265 3270 3275 3280 3285 3290 3295 3300 3305 3310 3315 3320 3325 3330 3335 3340 3345 3350 3355 3360 3365 3370 3375 3380 3385 3390 3395 3400 3405 3410 3415 3420 3425 3430 3435 3440 3445 3450 3455 3460 3465 3470 3475 3480 3485 3490 3495 3500 3505 3510 3515 3520 3525 3530 3535 3540 3545 3550 3555 3560 3565 3570 3575 3580 3585 3590 3595 3600 3605 3610 3615 3620 3625 3630 3635 3640 3645 3650 3655 3660 3665 3670 3675 3680 3685 3690 3695 3700 3705 3710 3715 3720 3725 3730 3735 3740 3745 3750 3755 3760 3765 3770 3775 3780 3785 3790 3795 3800 3805 3810 3815 3820 3825 3830 3835 3840 3845 3850 3855 3860 3865 3870 3875 3880 3885 3890 3895 3900 3905 3910 3915 3920 3925 3930 3935 3940 3945 3950 3955 3960 3965 3970 3975 3980 3985 3990 3995 4000 4005 4010 4015 4020 4025 4030 4035 4040 4045 4050 4055 4060 4065 4070 4075 4080 4085 4090 4095 4100 4105 4110 4115 4120 4125 4130 4135 4140 4145 4150 4155 4160 4165 4170 4175 4180 4185 4190 4195 4200 4205 4210 4215 4220 4225 4230 4235 4240 4245 4250 4255 4260 4265 4270 4275 4280 4285 4290 4295 4300 4305 4310 4315 4320 4325 4330 4335 4340 4345 4350 4355 4360 4365 4370 4375 4380 4385 4390 4395 4400 4405 4410 4415 4420 4425 4430 4435 4440 4445 4450 4455 4460 4465 4470 4475 4480 4485 4490 4495 4500 4505 4510 4515 4520 4525 4530 4535 4540 4545 4550 4555 4560 4565 4570 4575 4580 4585 4590 4595 4600 4605 4610 4615 4620 4625 4630 4635 4640 4645 4650 4655 4660 4665 4670 4675 4680 4685 4690 4695 4700 4705 4710 4715 4720 4725 4730 4735 4740 4745 4750 4755 4760 4765 4770 4775 4780 4785 4790 4795 4800 4805 4810 4815 4820 4825 4830 4835 4840 4845 4850 4855 4860 4865 4870 4875 4880 4885 4890 4895 4900 4905 4910 4915 4920 4925 4930 4935 4940 4945 4950 4955 4960 4965 4970 4975 4980 4985 4990 4995 5000 5005 5010 5015 5020 5025 5030 5035 5040 5045 5050 5055 5060 5065 5070 5075 5080 5085 5090 5095 5100 5105 5110 5115 5120 5125 5130 5135 5140 5145 5150 5155 5160 5165 5170 5175 5180 5185 5190 5195 5200 5205 5210 5215 5220 5225 5230 5235 5240 5245 5250 5255 5260 5265 5270 5275 5280 5285 5290 5295 5300 5305 5310 5315 5320 5325 5330 5335 5340 5345 5350 5355 5360 5365 5370 5375 5380 5385 5390 5395 5400 5405 5410 5415 5420 5425 5430 5435 5440 5445 5450 5455 5460 5465 5470 5475 5480 5485 5490 5495 5500 5505 5510 5515 5520 5525 5530 5535 5540 5545 5550 5555 5560 5565 5570 5575 5580 5585 5590 5595 5600 5605 5610 5615 5620 5625 5630 5635 5640 5645 5650 5655 5660 5665 5670 5675 5680 5685 5690 5695 5700 5705 5710 5715 5720 5725 5730 5735 5740 5745 5750 5755 5760 5765 5770 5775 5780 5785 5790 5795 5800 5805 5810 5815 5820 5825 5830 5835 5840 5845 5850 5855 5860 5865 5870 5875 5880 5885 5890 5895 5900 5905 5910 5915 5920 5925 5930 5935 5940 5945 5950 5955 5960 5965 5970 5975 5980 5985 5990 5995 6000 6005 6010 6015 6020 6025 6030 6035 6040 6045 6050 6055 6060 6065 6070 6075 6080 6085 6090 6095 6100 6105 6110 6115 6120 6125 6130 6135 6140 6145 6150 6155 6160 6165 6170 6175 6180 6185 6190 6195 6200 6205 6210 6215 6220 6225 6230 6235 6240 6245 6250 6255 6260 6265 6270 6275 6280 6285 6290 6295 6300 6305 6310 6315 6320 6325 6330 6335 6340 6345 6350 6355 6360 6365 6370 6375 6380 6385 6390 6395 6400 6405 6410 6415 6420 6425 6430 6435 6440 6445 6450 6455 6460 6465 6470 6475 6480 6485 6490 6495 6500 6505 6510 6515 6520 6525 6530 6535 6540 6545 6550 6555 6560 6565 6570 6575 6580 6585 6590 6595 6600 6605 6610 6615 6620 6625 6630 6635 6640 6645 6650 6655 6660 6665 6670 6675 6680 6685 6690 6695 6700 6705 6710 6715 6720 6725 6730 6735 6740 6745 6750 6755 6760 6765 6770 6775 6780 6785 6790 6795 6800 6805 6810 6815 6820 6825 6830 6835 6840 6845 6850 6855 6860 6865 6870 6875 6880 6885 6890 6895 6900 6905 6910 6915 6920 6925 6930 6935 6940 6945 6950 6955 6960 6965 6970 6975 6980 6985 6990 6995 7000 7005 7010 7015 7020 7025 7030 7035 7040 7045 7050 7055 7060 7065 7070 7075 7080 7085 7090 7095 7100 7105 7110 7115 7120 7125 7130 7135 7140 7145 7150 7155 7160 7165 7170 7175 7180 7185 7190 7195 7200 7205 7210 7215 7220 7225 7230 7235 7240 7245 7250 7255 7260 7265 7270 7275 7280 7285 7290 7295 7300 7305 7310 7315 7320 7325 7330 7335 7340 7345 7350 7355 7360 7365 7370 7375 7380 7385 7390 7395 7400 7405 7410 7415 7420 7425 7430 7435 7440 7445 7450 7455 7460 7465 7470 7475 7480 7485 7490 7495 7500 7505 7510 7515 7520 7525 7530 7535 7540 7545 7550 7555 7560 7565 7570 7575 7580 7585 7590 7595 7600 7605 7610 7615 7620 7625 7630 7635 7640 7645 7650 7655 7660 7665 7670 7675 7680 7685 7690 7695 7700 7705 7710 7715 7720 7725 7730 7735 7740 7745 7750 7755 7760 7765 7770 7775 7780 7785 7790 7795 7800 7805 7810 7815 7820 7825 7830 7835 7840 7845 7850 7855 7860 7865 7870 7875 7880 7885 7890 7895 7900 7905 7910 7915 7920 7925 7930 7935 7940 7945 7950 7955 7960 7965 7970 7975 7980 7985 7990 7995 8000 8005 8010 8015 8020 8025 8030 8035 8040 8045 8050 8055 8060 8065 8070 8075 8080 8085 8090 8095 8100 8105 8110 8115 8120 8125 8130 8135 8140 8145 8150 8155 8160 8165 8170 8175 8180 8185 8190 8195 8200 8205 8210 8215 8220 8225 8230 8235 8240 8245 8250 8255 8260 8265 8270 8275 8280 8285 8290 8295 8300 8305 8310 8315 8320 8325 8330 8335 8340 8345 8350 8355 8360 8365 8370 8375 8380 8385 8390 8395 8400 8405 8410 8415 8420 8425 8430 8435 8440 8445 8450 8455 8460 8465 8470 8475 8480 8485 8490 8495 8500 8505 8510 8515 8520 8525 8530 8535 8540 8545 8550 8555 8560 8565 8570 8575 8580 8585 8590 8595 8600 8605 8610 8615 8620 8625 8630 8635 8640 8645 8650 8655 8660 8665 8670 8675 8680 8685 8690 8695 8700 8705 8710 8715 8720 8725 8730 8735 8740 8745 8750 8755 8760 8765 8770 8775 8780 8785 8790 8795 8800 8805 8810 8815 8820 8825 8830 8835 8840 8845 8850 8855 8860 8865 8870 8875 8880 8885 8890 8895 8900 8905 8910 8915 8920 8925 8930 8935 8940 8945 8950 8955 8960 8965 8970 8975 8980 8985 8990 8995 9000 9005 9010 9015 9020 9025 9030 9035 9040 9045 9050 9055 9060 9065 9070 9075 9080 9085 9090 9095 9100 9105 9110 9115 9120 9125 9130 9135 9140 9145 9150 9155 9160 9165 9170 9175 9180 9185 9190 9195 9200 9205 9210 9215 9220 9225 9230 9235 9240 9245 9250 9255 9260 9265 9270 9275 9280 9285 9290 9295 9300 9305 9310 9315 9320 9325 9330 9335 9340 9345 9350 9355 9360 9365 9370 9375 9380 9385 9390 9395 9400 9405 9410 9415 9420 9425 9430 9435 9440 9445 9450 9455 9460 9465 9470 9475 9480 9485 9490 9495 9500 9505 9510 9515 9520 9525 9530 9535 9540 9545 9550 9555 9560 9565 9570 9575 9580 9585 9590 9595 9600 9605 9610 9615 9620 9625 9630 9635 9640 9645 9650 9655 9660 9665 9670 9675 9680 9685 9690 9695 9700 9705 9710 9715 9720 9725 9730 9735 9740 9745 9750 9755 9760 9765 9770 9775 9780 9785 9790 9795 9800 9805 9810 9815 9820 9825 9830 9835 9840 9845 9850 9855 9860 9865 9870 9875 9880 9885 9890 9895 9900 9905 9910 9915 9920 9925 9930 9935 9940 9945 9950 9955 9960 9965 9970 9975 9980 9985 9990 9995 10000 10005 10010 10015 10020 10025 10030 10035 10040 10045 10050 10055 10060 10065 10070 10075 10080 10085 10090 10095 10100 10105 10110 10115 10120 10125 10130 10135 10140 10145 10150 10155 10160 10165 10170 10175 10180 10185 10190 10195 10200 10205 10210 10215 10220 10225 10230 10235 10240 10245 10250 10255 10260 10265 10270 10275 10280 10285 10290 10295 10300 10305 10310 10315 10320 10325 10330 10335 10340 10345 10350 10355 10360 10365 10370 10375 10380 10385 10390 10395 10400 10405 10410 10415 10420 10425 10430 10435 10440 10445 10450 10455 10460 10465 10470 10475 10480 10485 10490 10495 10500 10505 10510 10515 10520 10525 10530 10535 10540 10545 10550 10555 10560 10565 1

- 49 -

Table 3. Magnesium concentration 2.0 mM

PCRmix	1x	100x
10xPCRbuffer (15 mM MgCl <sub>2</sub> )	5	500
MgCl <sub>2</sub> (25 mM)	1	100
dNTP (2.5 mM)	2.5	250
Q-solution	10	1000
Primer 4a (10 μM)	2	200
Primer 4b (10 μM)	2	200
Primer 8a (10 μM)	2	200
Primer 8b (10 μM)	2	200
TaqGold (5 units/μl)	0.25	25
H <sub>2</sub> O	13.25	1325
Sum:	40	4000
10 μl genomic DNA (2ng/μl)	was added to 40 μl PCR mix.	

PCR cycling conditions:

95°C 15min, 35x(94°C 30s, 55°C 1 min, 72°C 2 min), 72°C 10 min, 4°C.

20 Multiplex PCR using 6 amplification primers: A 50 μl PCR reaction was set up using Eu3, Eu6 and Eu10 SNP-specific fragments. All samples were amplified with the HotStarTaq Master Mix Kit from Qiagen adding Q-solution and MgCl<sub>2</sub> to a final concentration of 2.0 mM according to 25 the following protocol (Table 4).

Table 4. Magnesium concentration 2.0 mM

PCRmix	1x	100x
10xPCRbuffer	5	500
MgCl <sub>2</sub> (25 mM)	1	100
dNTP (2.5 mM)	2.5	250
Q-solution	10	1000
Primer 3a (10 μM)	2	200
Primer 3b (10 μM)	2	200
Primer 6a (10 μM)	2	200
Primer 6b (10 μM)	2	200
Primer 10a (10 μM)	2	200
Primer 10b (10 μM)	2	200
TaqGold (5 units/μl)	0.25	25
H <sub>2</sub> O	10.25	1025
Sum:	40	4500

10 μl genomic DNA (2ng/μl) was added to 40 μl PCR mix.

- 50 -

PCR cycling conditions:

95°C 15min, 35x(94°C 30s, 59°C 1 min, 72°C 2 min), 72°C 10 min, 4°C.

5

Sample Preparation

25 µl of PCR product (multiplex PCR product or pooled standard PCR product) was immobilised by the addition of 10 10 µl Dynabeads™ (Dynal Biotech ASA, supra) (10 µg/µl) together with 25 µl 2xBW buffer (10 mM Tris-HCl pH 7.57, 2M NaCl, 1 mM EDTA and 0.1% Tween 20). 15 pmol sequencing primer was added in annealing buffer (20 mM Tris-Acetate pH 7.51, 5 mM MgAc2) and the mixture 15 incubated for 2 minutes at 80°C. The samples were then allowed to cool to room temperature. 2.2 µg SSB (Amersham Pharmacia Biotech, supra) may be added at this point, if required.

20 Primer Extension

The primed DNA templates were placed in a microtiter plate containing Pyrosequencing™ substrates and enzymes (PSQ96™ plate, Pyrosequencing AB, supra). Nucleotides 25 were dispensed using fully automated microtiter-plate based PSQ™ Pyrosequencing™ instrument. The sequencing procedure was carried out by stepwise elongation of the primer-strand upon pre-specified addition of four different nucleotides. The templates were hybridized 30 with the extension primers mentioned above. The progress of sequencing was followed in real-time using Pyrosequencing™ software.

ResultsPrinciple of the SNP typing method.

5 The principle of the typing method described above is outlined in figure seven. In this model system, extension primers are hybridized to the target sample DNA, which is immobilised on magnetic beads.

10 In this set of experiments 3 sequencing primers for the RAAS system were used, either in isolation to show the 'simplex patterns' or in combination to show the multiplex patterns.

15 The signals resulting from the specific extension of each primer are directly correlated to the number of nucleotides incorporated. The 'fingerprint' produced can therefore be used to identify the genotype of the individual, against reference fingerprints, which can be theoretically deduced from the sequences of the variable regions. Reference fingerprints calculated

20 theoretically from the sequence of the SNPs are shown on Figure 8. 8a is the theoretical output for SNP Eu4 (ACE G2215A), 8b is the theoretical output for SNP Eu8 (ATG C521T) and 8c is the theoretical output for the simultaneous analysis of SNPs Eu4 and Eu8, the polymorphic positions are framed. 8d is the theoretical output for SNP Eu10 (ATP T573C), 8e is the theoretical output for SNP Eu6 (ACE T3409C), 8f is the theoretical

25 output for Eu3 (ACE T1237C) and 8g is the theoretical output for the simultaneous analysis of SNPs, Eu10, Eu6 and Eu3. SNPs Eu4, Eu10 and Eu6 are shown as heterozygotes, SNPs Eu8 as homozygote G and SNP Eu3 as homozygote A. These "reference" patterns can be used to

30 type the results shown in Figure 9: 9a is the sequencing data for SNP Eu4 (A/G), 9b is the sequencing data for SNP Eu8 (G/G) and 9c is the multiplex sequencing data

35

- 52 -

for the combination of SNP Eu4 (A/G) and SNP Eu8 (G/G), which correlates to theoretical output 8c, the frames indicating SNP positions. 9d, 9e and 9f are the sequencing data plots for SNP Eu10 (C/T), Eu6 (C/T) and Eu3 (A/A), respectively, and 9g is the multiplex sequencing data for the combination of these 3 SNPs, the polymorphic positions are boxed. Pattern 9g correlates to Figure 8g.

10 Example 3

Triplex genotyping on 4 SNPs in the RAAS System - Eu7 Eu8 (containing 2 SNPs), and Eu11.

15 Templates and Primers

PCR fragment	PCR primer 1 5'-biotin	PCR primer 2	From ref.
Eu7	TGA TGT AAC CCT CCT CTC CA	CGG CTT ACC TTC TGCTGT AGTA	US 619750 <sup>a</sup> ANPf4F
Eu8	CCA GGG CAG GGC TGA TA	CAA ACG GCT GCT TCA GGT	ANGe2f3F
Eu11	TTT CTC CTT CAA TTC TGA AAA GTA	GCC CCT CAG ATA ATG TAA GC	AT1-spec.1

25 Sequencing primers

Eu7 Eu7s 5' - ACG GCA GCT TCT TCC CC-3'

Eu8 A089RS (Eu8s) 5' - GCT GTG AAC ACG CCC AC-3'  
A060FS 5' - GCT GCT GCT CA-3'

30 Eu11 Eu11s 5' - GCA GCA CTT CAC TAC CAA AT-3'

PCR Amplification, sample preparation and primer extension reactions were performed as described in

- 53 -

Example 2, with the exception of Eu11, which was amplified according to the protocol in Table 2.

Table 2.

5

PCRmix	1x	100x
10xPCRbuffer	5	500
MgCl <sub>2</sub> (25 mM)	3	300
dNTP (2.5 mM)	2.5	250
DMSO	0	0
Primer 11a (10 μM)	2	200
Primer 11b (10 μM)	2	200
TaqGold (5 units/μl)	0.3	30
H <sub>2</sub> O	30.2	3020
Sum:	45	4500

5 μl genomic DNA (2ng/μl) was added to 45 μl PCR mix.

PCR cycling conditions for Eu11:

20

95°C 5 min, 50x(95°C 315s, 52°C 30s, 72°C 45s), 72°C 5 min, 4°C.

Results

25

In this set of experiments 3 sequencing ("extension") primers for the RAAS system were used, and the signals resulting from the specific extension of each primer can be directly correlated to the number of nucleotides incorporated. Theoretical reference patterns are shown in Figure 10, which can be used to determine the genotype shown in figure 11. 10a, 10b and 10c are the theoretical outputs obtained for SNPs Eu8 (G/G), Eu7 (C/C and T/T) and Eu11 (A/G), with the theoretical multiplex output shown on figure 10d. This correlates to the actual results obtained shown in figure 11. The polymorphic positions are boxed, and the genotype of this individual is Eu8 G/G, Eu7 C/C and T/T and Eu11 A/G. The pyrogram exhibits some nucleotide background incorporation which can be reduced as discussed

40

- 54 -

previously (e.g. add SSB after primer annealing).

Example 4

5 SNP typing in Human Coagulation Factor V, Prothrombin and Plasminogen activator inhibitor.

**Introduction**

10 Thrombosis is a complex (multifactorial) trait. The genes involved are typically susceptibility genes, where the differences are not point mutations but particular forms (alleles) of polymorphisms. The disorder results from the presence of an increased frequency of specific 15 alleles in unfavorable combinations.

20 During the last ten to fifteen years, mutation or variation in several genes has been found to be associated with venous thrombosis. This includes genes such as factor V (FV), prothrombin (FII) and plasminogen activator inhibitor (PAI1).

25 Coagulation Factor V (FV) and Prothrombin (FII) are both essential components in the human coagulation cascade, which ultimately results in the stemming of blood loss. Prothrombin is proteolytically cleaved in the first step of this cascade converting into the clotting enzyme thrombin. Coagulation factor V serves as a cofactor for the coagulation factor X-catalyzed activation of 30 prothrombin to thrombin. Point mutations in these genes may cause impairments in processes of thrombosis and hemostasis. One such is venous thrombosis, predominantly afflicting people of European origin. The mutations, Factor V Leiden (FV:G1691A) and the G20210-A 35 prothrombin variant (FII:G20210A), are the two single most important genetic risk factors for developing venous thrombosis. This European predisposition has

- 55 -

been explained to some extent by the characterization by these two variants. In addition to these two established risk factors for venous thrombosis, the role of other genetic variations is still under investigation 5 (Martnelli et al., 1998; De Stefano et al., 1999; Rees et al., 1999; Hessner et al., 1999).

Several prospective studies have documented that the fibrinolytic capacity is an important determinant of the 10 risk of thrombosis. Many studies have convincingly shown that survivors of myocardial infarction have impaired fibrinolytic activity because of increased concentrations of plasma plasminogen activator inhibitor-1 (PAI-1). A single guanosine 15 insertion/deletion polymorphism in the promoter region of the PAI1 gene, commonly called 4G/5G, has been shown to be associated with plasma PAI-1 activity (Dawson et al., 1993; Eriksson et al., 1995).

20 **Primers**

Three sets of PCR primers were designed. The fragment spanning over exon 10 and intron 10 of human coagulation factor V was 162 bp long, the prothrombin fragment 25 spanning over exon 14 and intron 14 was 211 bp and the fragment in the promotor region of the PAI1 gene was 152 bp. One primer in each set was biotinylated in order to allow subsequent immobilization to magnetic beads/sepharose beads. In addition, three sequencing 30 primers were designed to hybridize in close proximity to the factor V Leiden SNP, the G20210A prothrombin variant and the 4G/5G deletion of PAI1 see Figure 12.

**PCR primers:**

35 Prothrombin

A001FPB Biotin-5'-CCT GAA GAA GTG GAT ACA GAA GG-3'

A008RP 5'-CAG TAG TAT TAC TGG CTC TTC CTG A-3'

- 56 -

Factor V PSO90 5'-GGG CTA ATA GGA CTA CTT CTA ATC-3'  
PSO91B Biotin-5'-TCT CTT GAA GGA AAT GCC CCA TTA-3'

PAI1 PSO112FPB Biotin-5'- CCC ACC CAG CAC ACC TC-3'  
5 PSO113RP 5'- GAC TCT TGG TCT TTC CCT CAT C-3'

Sequencing primers:

Prothrombin A009SR 5'-ACT GGG AGC ATT GAG-3'  
Factor V PSO83 5'-AGC AGA TCC CTG GAC-3'  
10 PAI1 A114SR 5'- CAC GGC TGA CTC CCC-3'

PCR amplification

A 50  $\mu$ l PCR reaction was set up using HotStarTaq Master  
15 Mix Kit from QiaGen according to the following protocol  
(Table 5).

**Table 5.** Magnesium concentration 2.0 mM

PCRmix	1x	100x
20 10xPCRbuffer (15 mM MgCl <sub>2</sub> )	5	500
MgCl <sub>2</sub> (25 mM)	4	400
dNTP (2.5 mM)	2.5	250
A001FPB (10 mM)	1	100
A008RP (10 mM)	1	100
25 PSO90 (10 mM)	1	100
PSO91B (10 mM)	1	100
PSO112FPB (10 mM)	1	100
PSO113RP (10 mM)	1	100
HotStarTaq (5 units/ml)	0.2	20
30 H <sub>2</sub> O	29.3	2930
Sum:	45	4500

5  $\mu$ l genomic DNA (2ng/ $\mu$ l) was added to 45  $\mu$ l PCR mix.

PCR cycling conditions:

35 95°C 5min, 50x(95°C 30s, 67°C 45s, 72°C 60s), 72°C 5min,  
4°C.

Sample preparation and primer extension

40 Were performed as described in Example 2.

Results

The theoretical output obtained by typing each SNP or deletion individually are shown as figures 13a, 13b and 5 13c, representing PAI1 genotype for 4G/5G deletion (C/C), SNP G20210A prothrombin (C/T) and SNP G1691A Factor V Leiden (A/G), respectively. The theoretical 10 multiplexing output for the multiplex assay of these 3 SNPs is shown as figure 13d, with the deletion or SNP position shown. Figures 13e to 13j represent the theoretical 15 output expected for 6 genotypes upon which real data was then collected, see Figure 14. The pyrograms shown in figure 14 are 6 possible genotypes that can be present in the human population in these 20 genes. 14a is the results from the genotype PAI1 C/C, prothrombin C/C and factor V G/G, 14b is the genotype PAI1 del/del, Prothrombin C/C and factor V A/A, 14c is the genotype PAI1 del/C, Prothrombin T/T and Factor V G/G, 14d is the genotype PAI1 C/C, prothrombin T/C and 25 Factor V G/G, 14e is the genotype PAI1 C/del, Prothrombin C/C, Factor V G/G and 14f is the genotype PAI1 C/C, prothrombin C/C and Factor V A/G. Figure 14a corresponds to 13e, 14b to 13f, 14c to 13g, 14d to 13h, 14e to 13i and 14f to 13j.

25

Example 5

## CYP2D6 SNP analysis

30

Introduction

The CYP2D6 gene is a member of the cytochrome P450 gene superfamily, which in total consists of nine gene families. Four of these gene families are responsible 35 for the metabolism and elimination of most foreign chemicals that enters the body via ingestion. The human CYP2D locus is mapped to chromosome 22q13.1 (Gough

- 58 -

et.al, 1993). The CYP2D6 gene encodes for an enzyme, debrisoquine 4-hydroxylase, which is involved in the metabolism of more than 40 drugs, among them neuroleptics, antidepressants, antiarrhythmics, 5 b-blockers and opioids. The enzyme is characterised by extreme variability in activity (interindividual and interethnic). The CYP2D6 genotype and catalytic function are closely coupled, and genotyping could be an important tool for determining drug doses for 10 individuals. More than 50 alleles have been identified, of which many encodes for a non-functional enzyme. The alleles are defined by a number of variations; SNPs, insertion or deletions of single base pairs, deletion of the complete gene, and duplications of the gene. The 15 sequences analysed in this example are as follows:

G1846T: GCCAACCACTCC G/T GT  
G1934A: G/A GACGCCCTTCG  
T1795del: GCAG(T)GGGTGACCG  
20 G1749C: G/C CTCCACCTTGCG

Primers

Table 6. PCR primers and sequencing primers in the 25 multiplex method. The primers are named F for a forward direction and R for a reversed direction. P represents a PCR primer, and S a sequencing primer and B means biotin labelled in the 5'end.

Fragment	Primers	Primer sequence	Sequence to be identified
5	A061RPB	B-CCTCGGTCTCTCGCTCCGC	
	A118FP	GAGCAGAGGCGCTTCTCCGT	
	A143FS	CCTTCGCCAACAC	TCCG/TGT
	A182FS	CAAGAACGCTGGAG	CAG (T) GGGTG
	A183FS	GCATCTCCCACCCCC	AG/AGACGCCCTTC
10	2162	B-ACTGTTCCCAGATGGCTC	
	A062FP	GACCCCGTTCTGTCTGGTGT	
	A145FS	TTCAATGATGAGAAC	TGC/TG
	A146FS	CCTGCTCATGATCCT	ACA/CTCCGG
	A147FS	TGAGCTGCTAACTGA	GCAC (A) GG

Figure 15 shows the localisation of the primers in the CYP2D6 nucleic acid fragments for the multiplex method:  
15 fragments 2162 and 61118.

#### PCR amplification

A nested PCR amplification was performed. For both the  
20 first and the nested 50 µl PCR reaction HotStarTaq Master Mix Kit from QiaGen was used and was set up according to the following protocol (Table 7).

**Table 7.** Magnesium concentration 1.5 mM

PCRmix	1x	100x
10xPCRbuffer (15 mM MgCl <sub>2</sub> )	5	500
MgCl <sub>2</sub> (25 mM)	0	0
dNTP (2.5 mM)	4	400
Primer 1 (10 mM)	1	100
Primer 2 (10 mM)	1	100
HotStarTaq (5 units/µl)	0.5	50
H <sub>2</sub> O	37.5	3750
Sum:	49	4900

35 PCR 1.

1 µl genomic DNA (10 ng/µl) was added to 49 µl PCR mix.  
PCR cycling conditions:

- 60 -

PCR method, primary PCR (fragment 4142)  
95°C 15min, 25x(95°C 45s, 66°C 45s, 72°C 60s), 72°C  
5min, 4°C

5 PCR method, secondary PCR  
95°C 5min, 20x(95°C 45s, T<sub>A</sub> 45s, 72°C 45s), 72°C 5min,  
4°C

10 T<sub>A</sub>, fragment 61118 was 61°C  
2162 was 63°C

#### Sample preparation

15 Took place as described in example 2. SSB was added to  
the primer/template mix after hybridisation. 0,55 µg  
SSB was added for fragment 2162 and 2,2 µg for fragment  
61118. The amounts of sequencing primers were for  
fragment 2162: 15 pmoles of each, and for fragment  
61118: 5 pmoles of primers 182 and 183, and 70 pmoles of  
20 primer 143.

#### Primer extension

Was performed as described for example 2.

25

#### Results

30 Two theoretical output for fragment 61118 in a multiplex  
analysis are shown as figures 16a and 16b. 16a shows  
the genotype G<sub>1934</sub>A (A/G), G<sub>1749</sub>C (C/G), T<sub>1795</sub>del (no  
deletion) and G<sub>1846</sub>T (T/G) and figure 16b differs in that  
T<sub>1795</sub>del shows the deletion of the T residue.

35 Figure 17 shows the actual results from genotype  
established from the pyrogram is G<sub>1934</sub>A (G/G), G<sub>1749</sub>C  
(G/G), T<sub>1795</sub>del (no deletion ∴ T/T) and G<sub>1846</sub>T (G/G). This  
is a different genotype to those shown in figure 16.

- 61 -

This demonstrates that it is possible to type multiple SNPs and deletions on one fragment of nucleic acid using multiple extension primers.

5       EXAMPLE 6

Serum samples

72 sera from HCV-positive Veterans were obtained from  
10 Stanford Veteran hospital. Five HCV-positive sera were  
obtained from Iran.

Synthesis and purification of oligonucleotides

15       The oligonucleotides HCV-PCR-OUTF  
(5'- CCCTGTGAGGAACTWCTGTCTTCACGC) , HCV-PCR-OUTR  
(5'-GCTCATGRTGCACGGTCTACGAGACCT) , HCV-PCR-INF  
(5'-TCTAGCCATGGCGTTAGTAYGAGTGT) , BHCV-PCR-INR  
(5'-Biotin-CACTCGCAAGCACCCATTAGGCAGT) , HCV-SEQF1  
20       (5'-GGAACCGGTGAGTACACCGGAAT) , HCV-SEQF2  
(5'-GACYGGGTCTTCTTGGA) , HCV-SEQF3  
(5'-ATTTGGGCGTGCCCCCGC) , were all synthesized and HPLC  
purified by MWG Biotech (High points, NC, USA) .

25       RNA extraction, cDNA synthesis and amplification

RNA was extracted from 50  $\mu$ l of serum. cDNA was  
synthesized using AMV reverse transcriptase on HCV cDNA  
obtained from different patients using BHCV-PCR-INR and  
30       HCV-PCR-INF to generate a 270 base long product.

The biotinylated PCR products were immobilized onto  
streptavidin-coated super paramagnetic beads Dynabeads<sup>TM</sup>  
M280-Streptavidin (Dynal A.S., Oslo, Norway). Single-  
35       stranded DNA was obtained by removing the supernatant  
after incubation of the immobilized PCR product in 0.10  
M NaOH for 3 min. Five pmol of sequencing primers HCV-

- 62 -

SEQF1, HCV-SEQF2, and HCV-SEQF3 were hybridized to the immobilized strand.

Primer extension reaction

5

The primed DNA template were placed in a microtiter plate containing 0.5 µg SSB (Amersham Pharmacia Biotech, USA), and Pyrosequencing™ substrates (www.pyrosequencing.com Pyrosequencing AB, Uppsala, Sweden) and enzymes were dispensed using fully automated microtiter plate-based PSQ™ Pyrosequencing™ instrument. The sequencing procedure was carried out by stepwise elongation of the primer-strand upon pre-specified addition of four different nucleotides. The template was hybridized with the three extension primers described above. The progress of sequencing was followed in real-time using Pyrosequencing™ SNP software, (Pyrosequencing™ AB, Uppsala, Sweden) and subtyping was performed manually.

10  
15  
20

Results

**Principle of the typing method.**

25

The principle of the typing method described above is outlined in Figure 1. In this model system, extension primers are hybridized to the target sample DNA, which is immobilized on magnetic beads.

30

The extension primers hybridise specifically to the conserved region adjacent to the variable region.

35

The signals resulting from the specific extension of each primer are directly correlated to the number of nucleotides incorporated. The 'fingerprint' produced can therefore be used to identify the genotype of the individual, against reference fingerprints, which can be

- 63 -

theoretically deduced from the sequences of the variable regions. References fingerprints calculated theoretically from the sequence of the variable regions are shown on Figure 3. These can be used to type the 5 results shown on Figure 4: Figure 4a is the fingerprint for HCV 1a, Figure 4b is the fingerprint for HCV 1b, Figure 4c is the fingerprint for HCV 2a, Figure 4d is the fingerprint for HCV 2b, Figure 4e is the fingerprint for HCV 3a, and Figure 4f is the fingerprint for HCV 3b. 10 Therefore, using the method of the invention, it was possible to genotype HCV infection. Of the 77 sera analyzed by the method of the invention. 35% were infected with HCV 1a, 29% with HCV 1b, 21% with HCV 2a, 4% with HCV 2b, 10% with HCV 3a and 1% with HCV 3b.

Claims

1. A method of typing one or more nucleic acid molecules, said method comprising:

5       simultaneously or sequentially performing two or more primer extension reactions, each primer binding at a different predetermined site in said nucleic acid molecule(s), and determining the pattern of nucleotide incorporation to obtain a test pattern for said nucleic acid molecule(s) which is optionally compared with one or more reference patterns to type the said nucleic acid molecule(s).

15      2. A method as claimed in claim 1 wherein the nucleic acid contains two or more variable sites.

20      3. A method for obtaining typing information about a plurality of variable sites within target nucleic acid, comprising simultaneously or sequentially performing two or more primer extension reactions, each primer binding at a different predetermined site in said target nucleic acid, the pattern of nucleotide incorporation providing the typing information about said variable sites.

25      4. A method as claimed in any one of claims 1 to 3 wherein the pattern of nucleotide incorporation is determined by sequencing-by-synthesis.

30      5. A method as claimed in any one of claims 1 to 4 wherein nucleotides are added to the reaction mix sequentially in a predetermined order.

35      6. A method as claimed in any one of claims 1 to 5 wherein nucleotide incorporation is determined quantitatively.

7. A method as claimed in any one of claims 2 to 6

- 65 -

wherein if nucleotide incorporation takes place at one variable site, there is no nucleotide incorporation at the other variable site(s).

5 8. A method as claimed in any one of claims 2 to 7 wherein a first extension primer binds closer to its variable site than a second primer does to its variable site.

10 9. A method as claimed in claim 8 wherein the second primer is 10-20 nucleotides further away from its variable site than is said first primer.

15 10. A method as claimed in any one of claims 1 to 9 wherein single-stranded binding protein is added to the reaction mix after the primers are annealed to the nucleic acid template.

20 11. A method as claimed in any one of claims 1 to 10 wherein the primer extension reactions occur simultaneously.

12. A method as claimed in any one of claims 2 to 11 wherein 3 or more variable sites are typed.

25 13. A method as claimed in any one of claims 2 to 11 wherein 3 or more primer extension reactions are performed.

30 14. A method of diagnosis of pathological conditions characterised by the presence of specific nucleic acid molecule(s), comprising simultaneously or sequentially performing two or more primer extension reactions, each primer binding at a different predetermined site in said nucleic acid molecule(s), the pattern of nucleotide incorporation allowing diagnosis of said pathological conditions.

- 66 -

15. A kit for use in a method of typing nucleic acid which comprises:

optionally primer(s) for *in vitro* amplification; two or more primers for primer extension reactions; 5 nucleotides for amplification and/or for the primer extension reaction; optionally a polymerase enzyme for the amplification and/or primer extension reaction; and optionally means for detecting primer extension

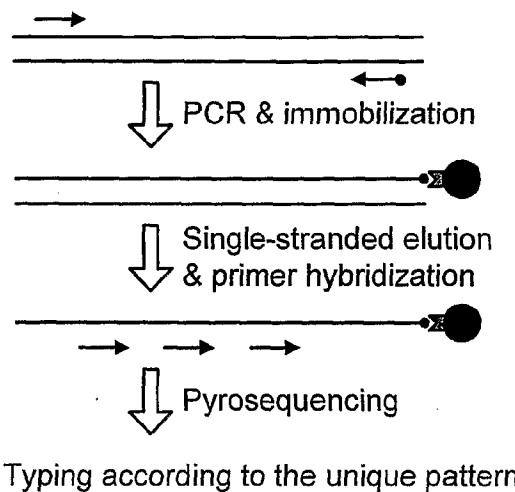


FIG. 1

## A Primer position: 3'-end 4 bp from SNP position

-----**CCGGTGCAGGAGACGAGACGAGGTCCATGAAG/ACAGTC**-----  
 -----**CTCTGCTCCAGGTACTTC/TGTCA**-----

## B Primer position: 3'-end 5 bp from SNP position

-----**CCGGTGCAGGAGACGAGACGAGGTCCATGAAG/ACAGTC**-----  
 -----**GCTCTGCTCCAGGTACTTC/TGTCA**-----

## C Primer position: 3'-end 10 bp from SNP position

-----**CCGGTGCAGGAGACGAGACGAGGTCCATGAAG/ACAGTC**-----  
 -----**CTCGCTCTGCTCAGGTACTTC/TGTCA**-----

FIG. 5

**SUBSTITUTE SHEET (RULE 26)**

FIG. 2

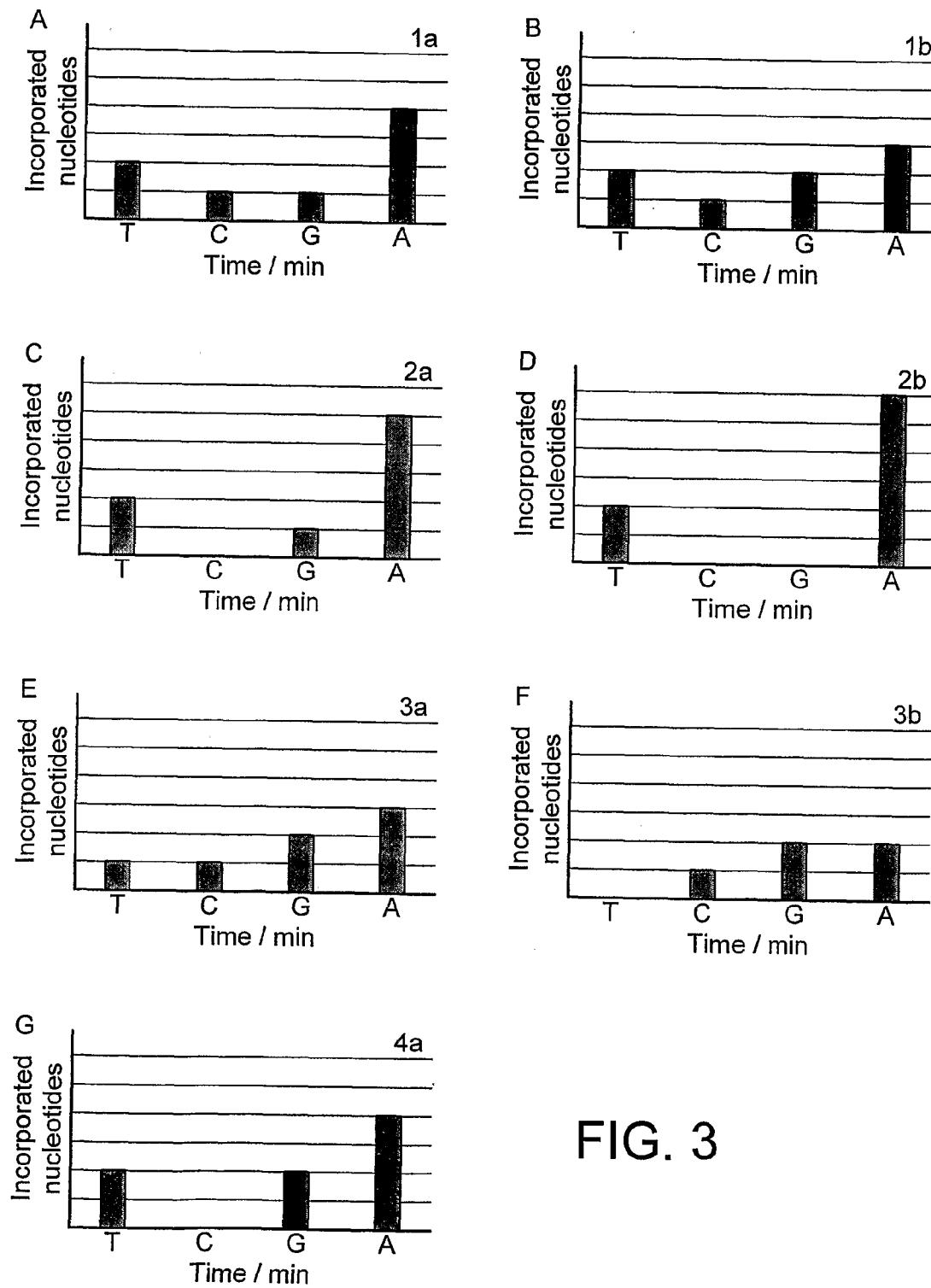


FIG. 3

4 / 19

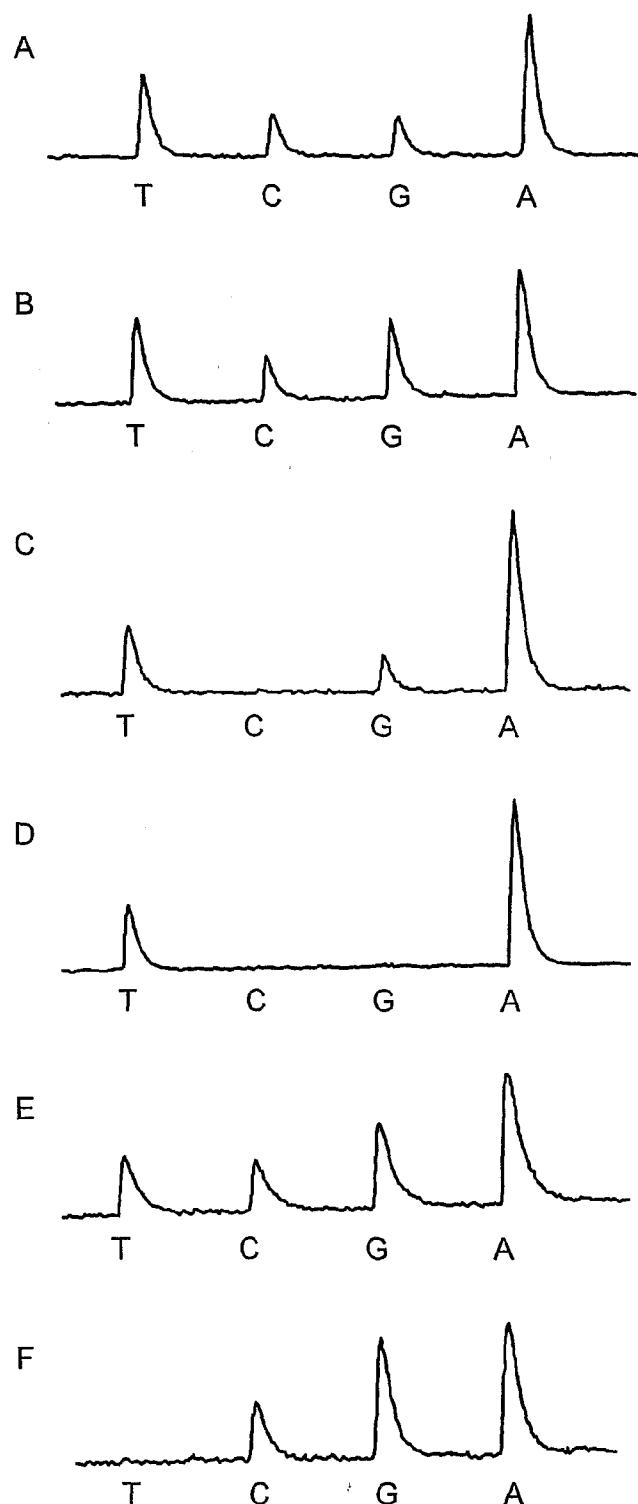


FIG. 4

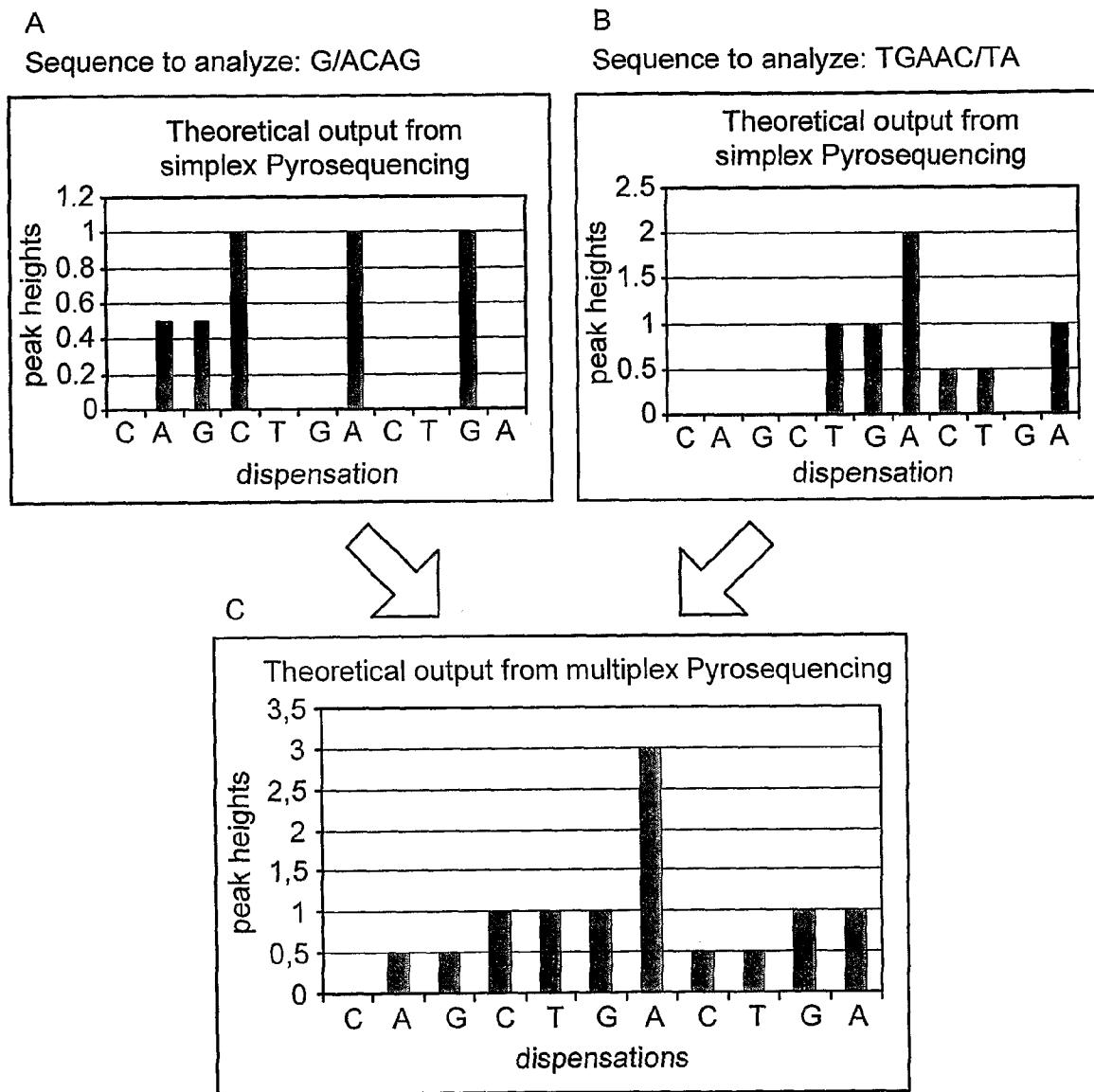


FIG. 6

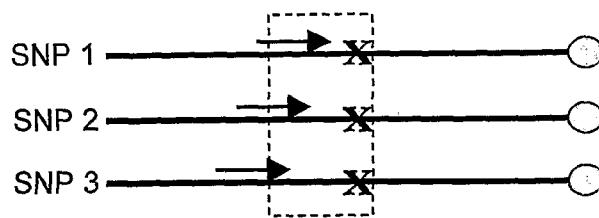


FIG. 7

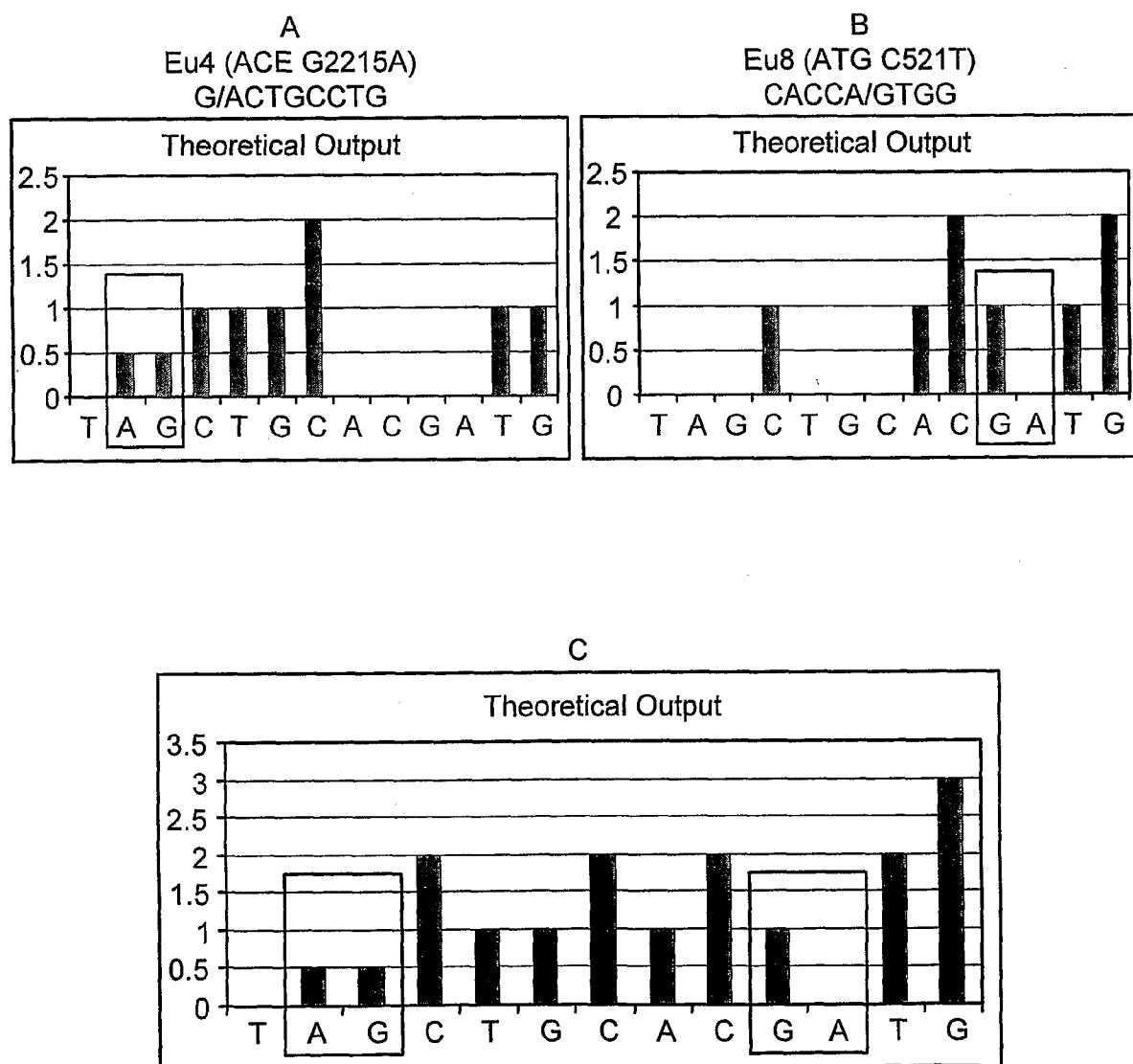
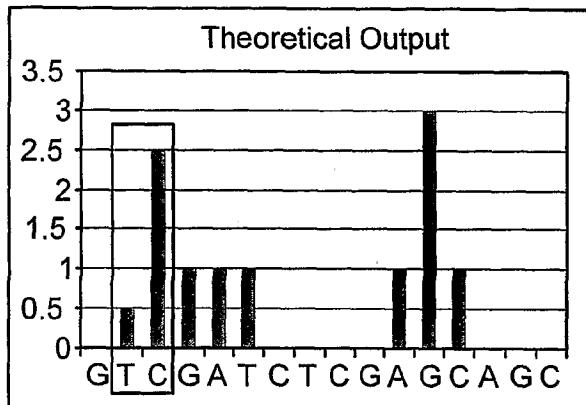
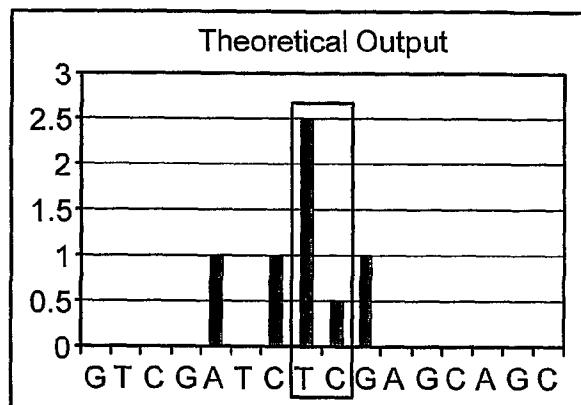


FIG. 8

D  
Eu10 (ATP T573C)  
C/TCCGATAGGGC



E  
Eu6 (ACE T3409C)  
ACTTC/TG



F  
Eu3 (ACE T1237C)  
AGACA/GGGC

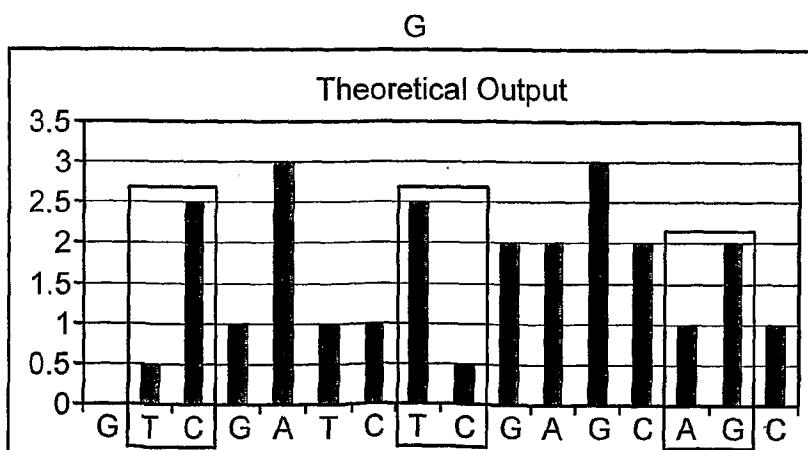
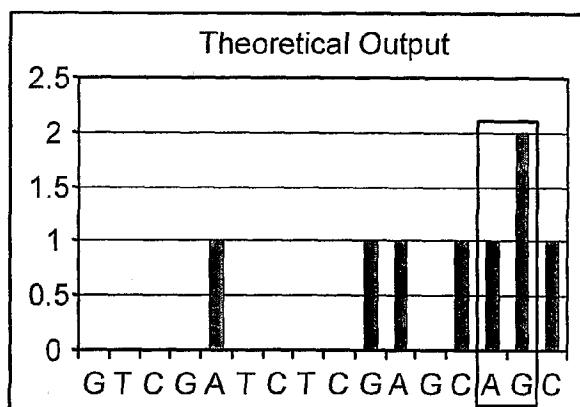


FIG. 8 CONT'D

8 / 19

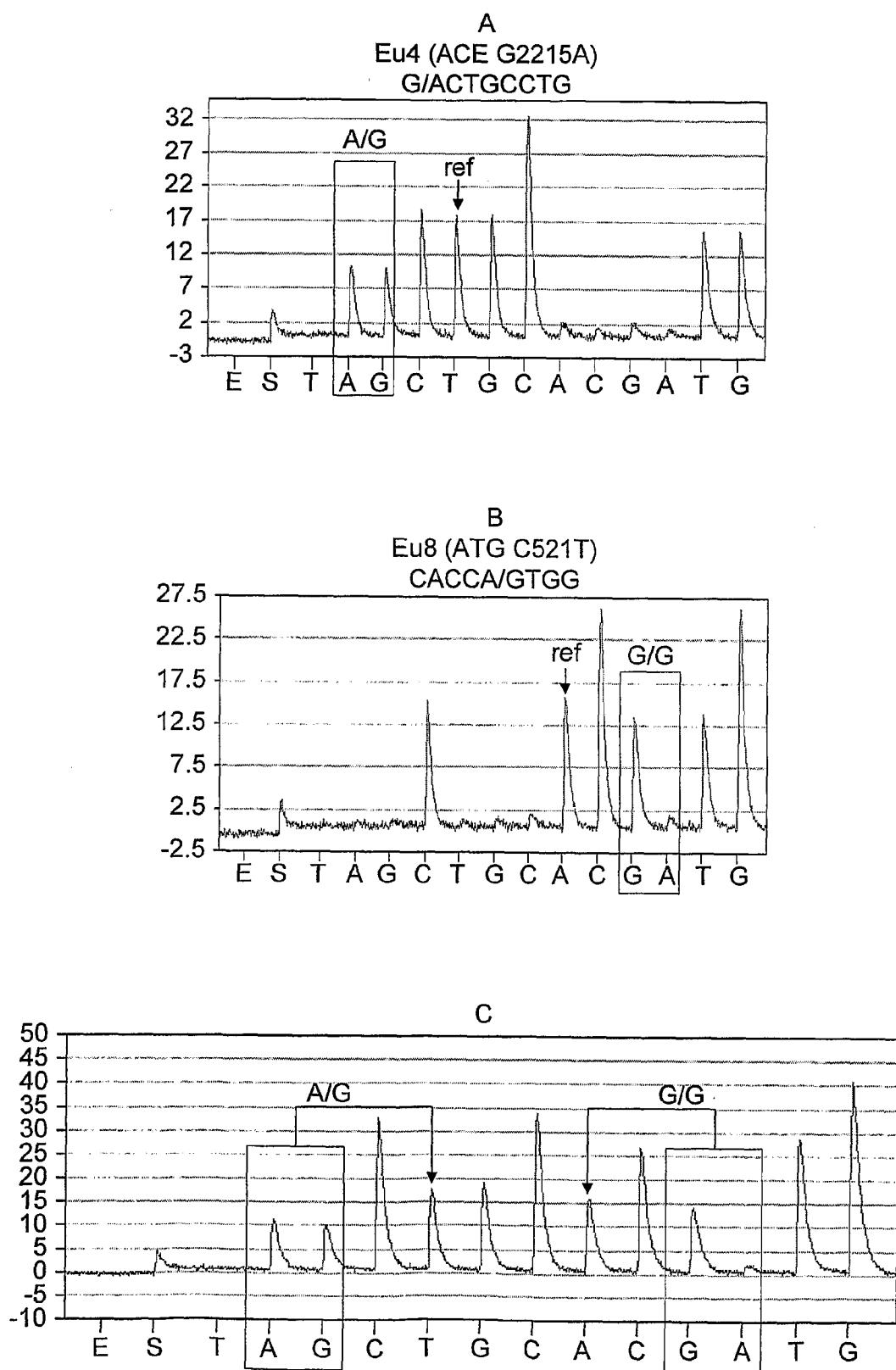


FIG. 9

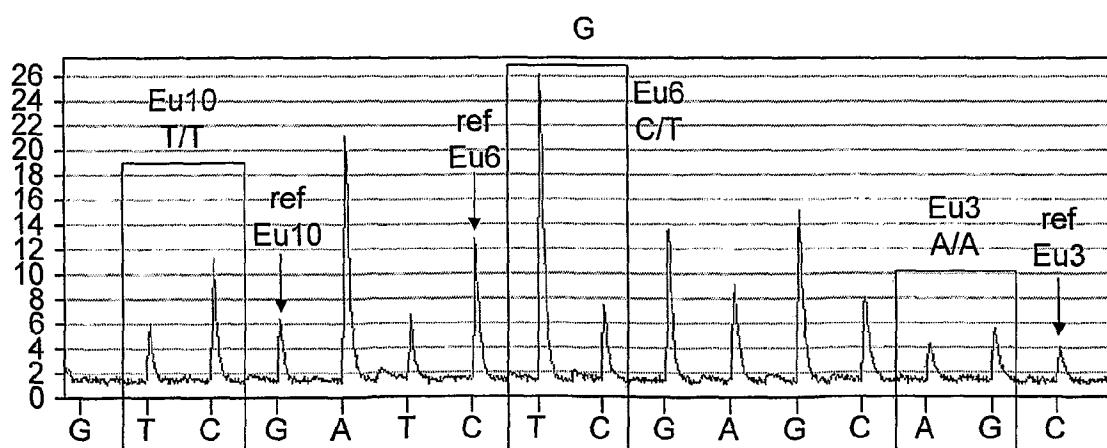
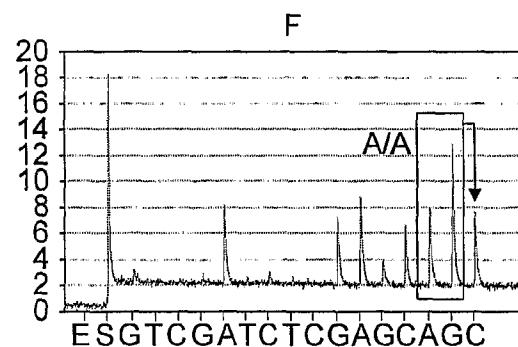
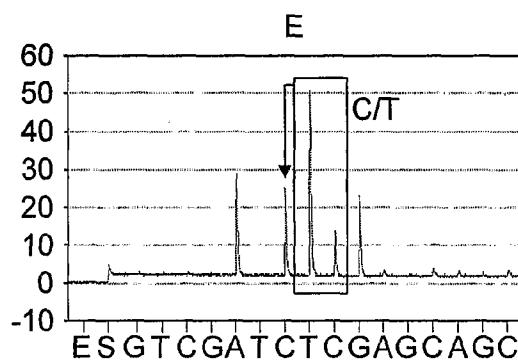
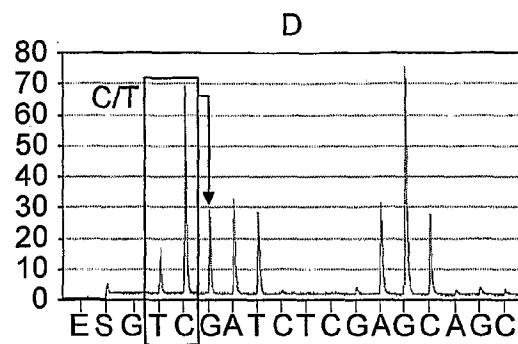


FIG. 9 CONT'D

10 / 19

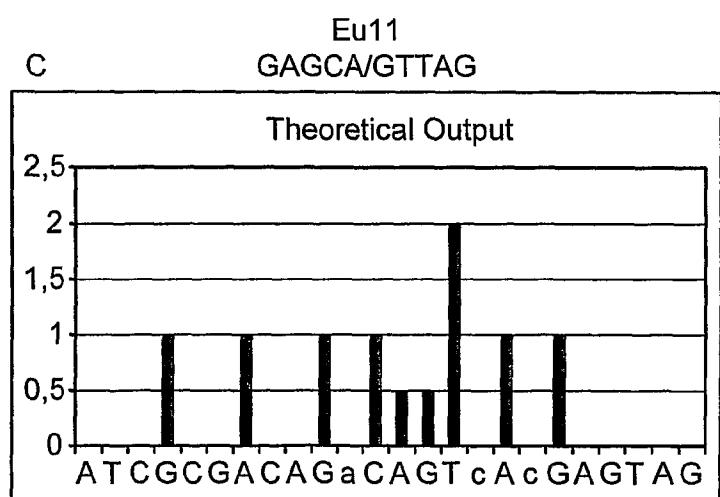
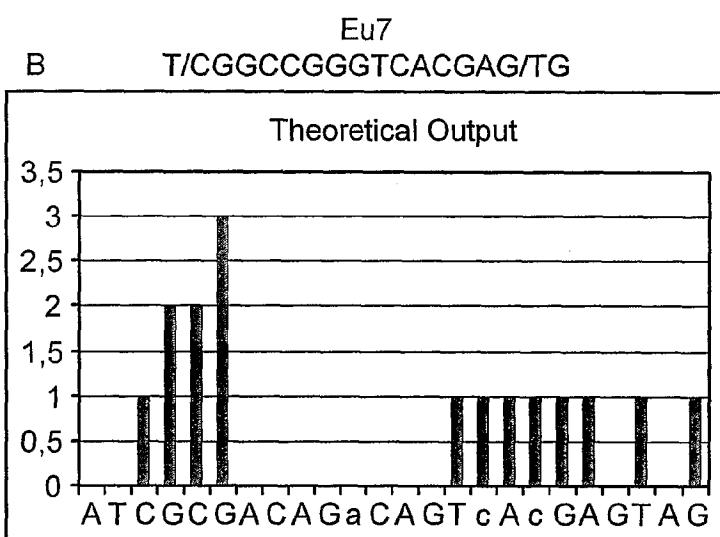
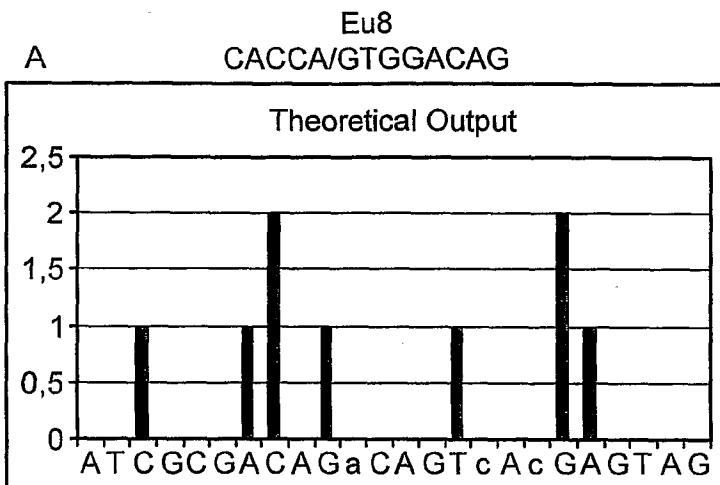


FIG. 10

11 / 19

D

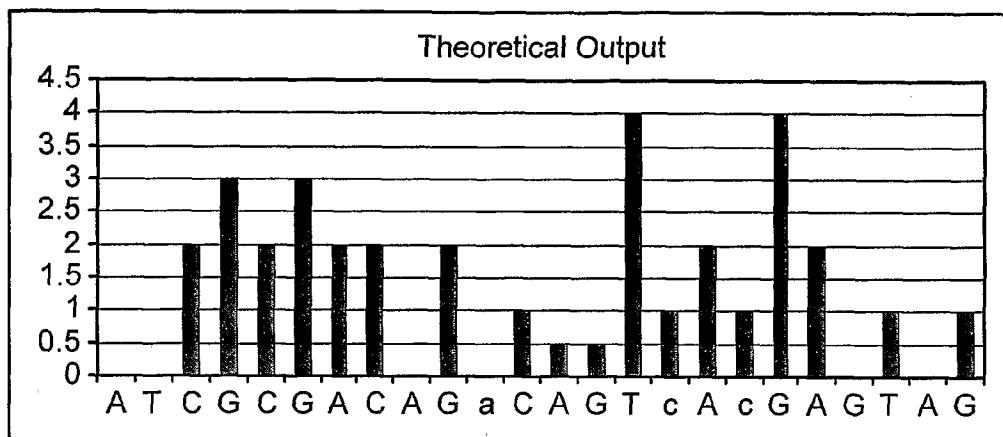


FIG. 10 CONT'D

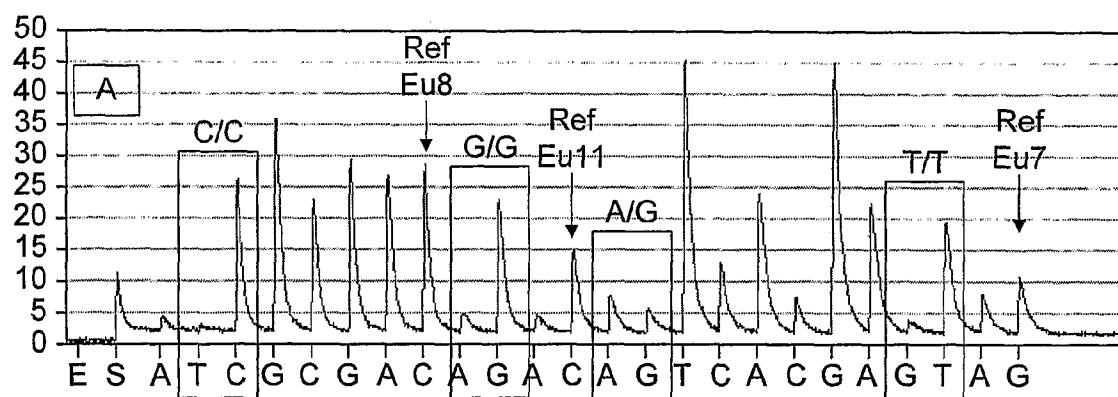
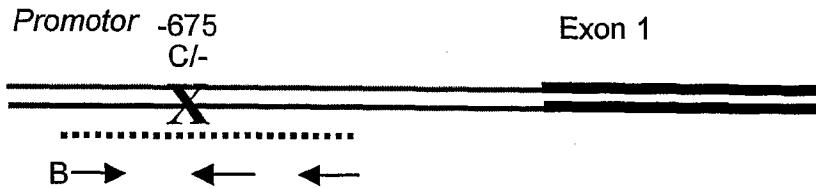


FIG. 11

12 / 19

## A. Plasminogen Activator Inhibitor I



## B. Prothrombin



## C. Factor V



FIG. 12

A Plasminogen-activator inhibitor 1: (C)ACGTG

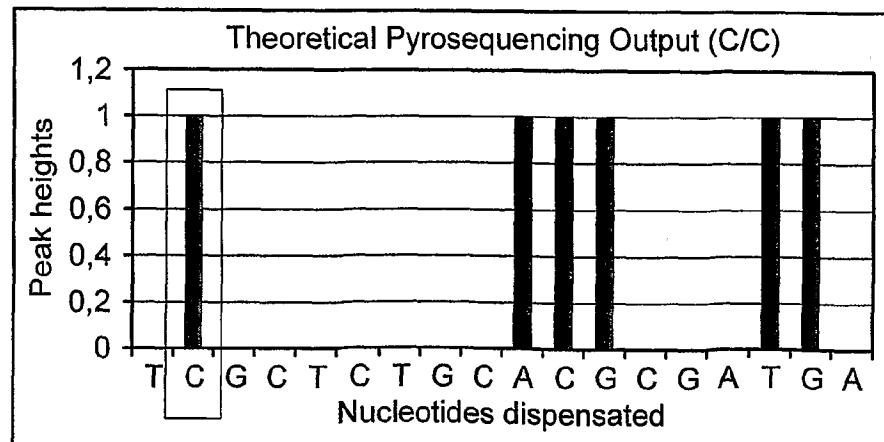


FIG. 13

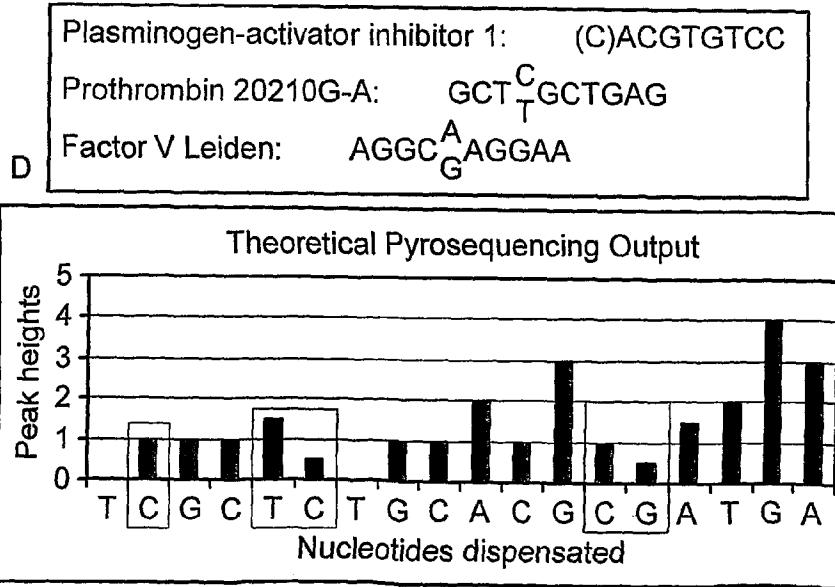
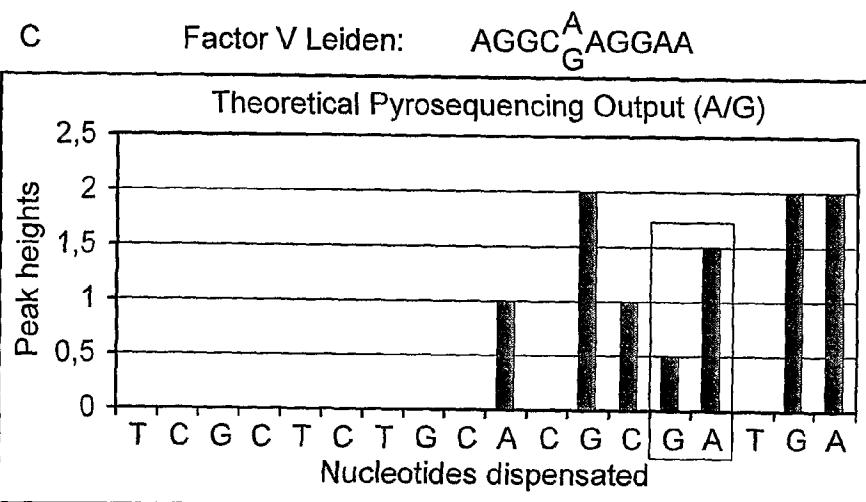
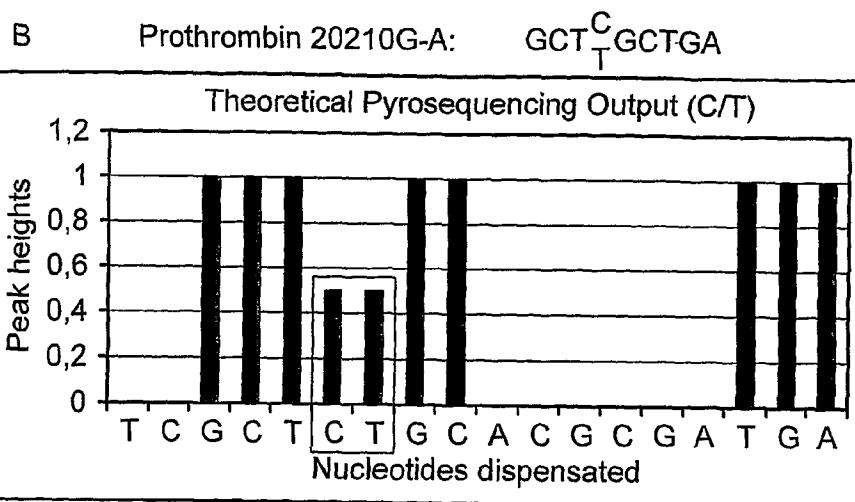


FIG. 13 CONT'D

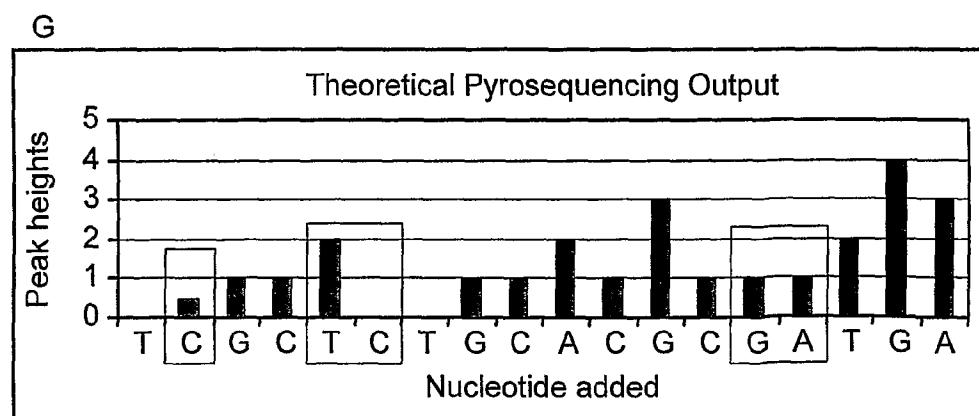
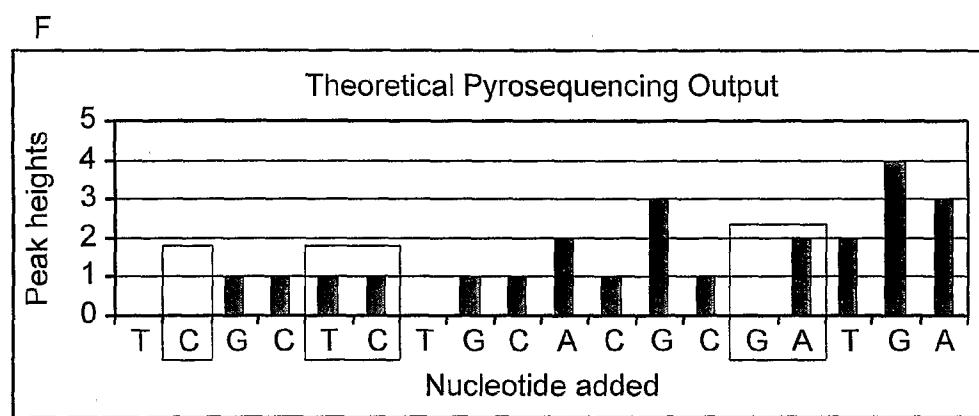
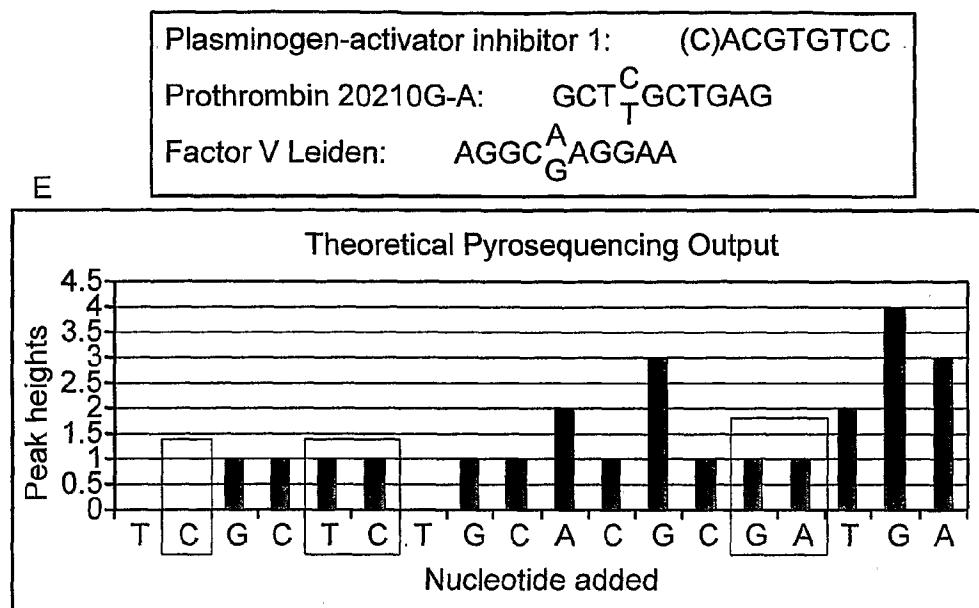
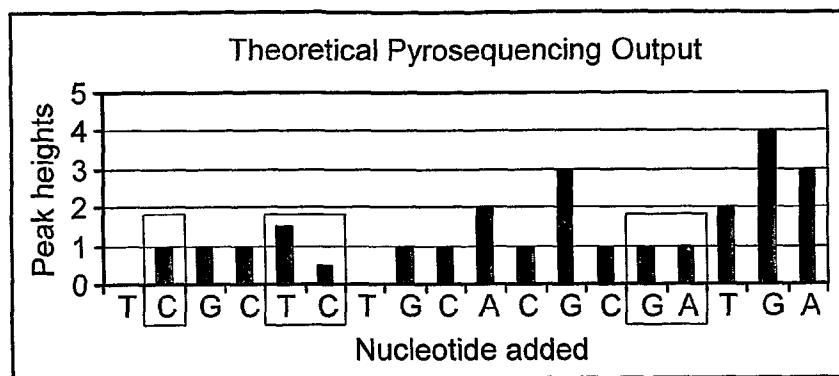


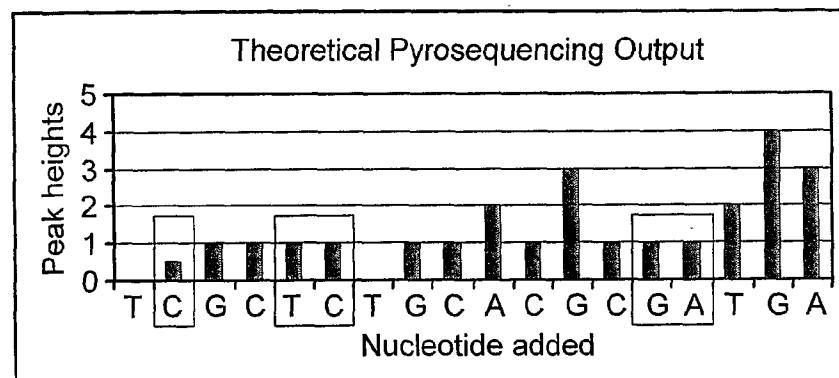
FIG. 13 CONT'D

15 / 19

H



I



J

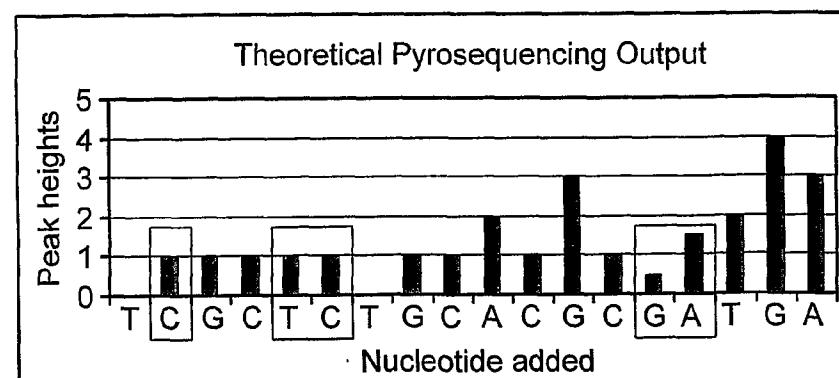


FIG. 13 CONT'D

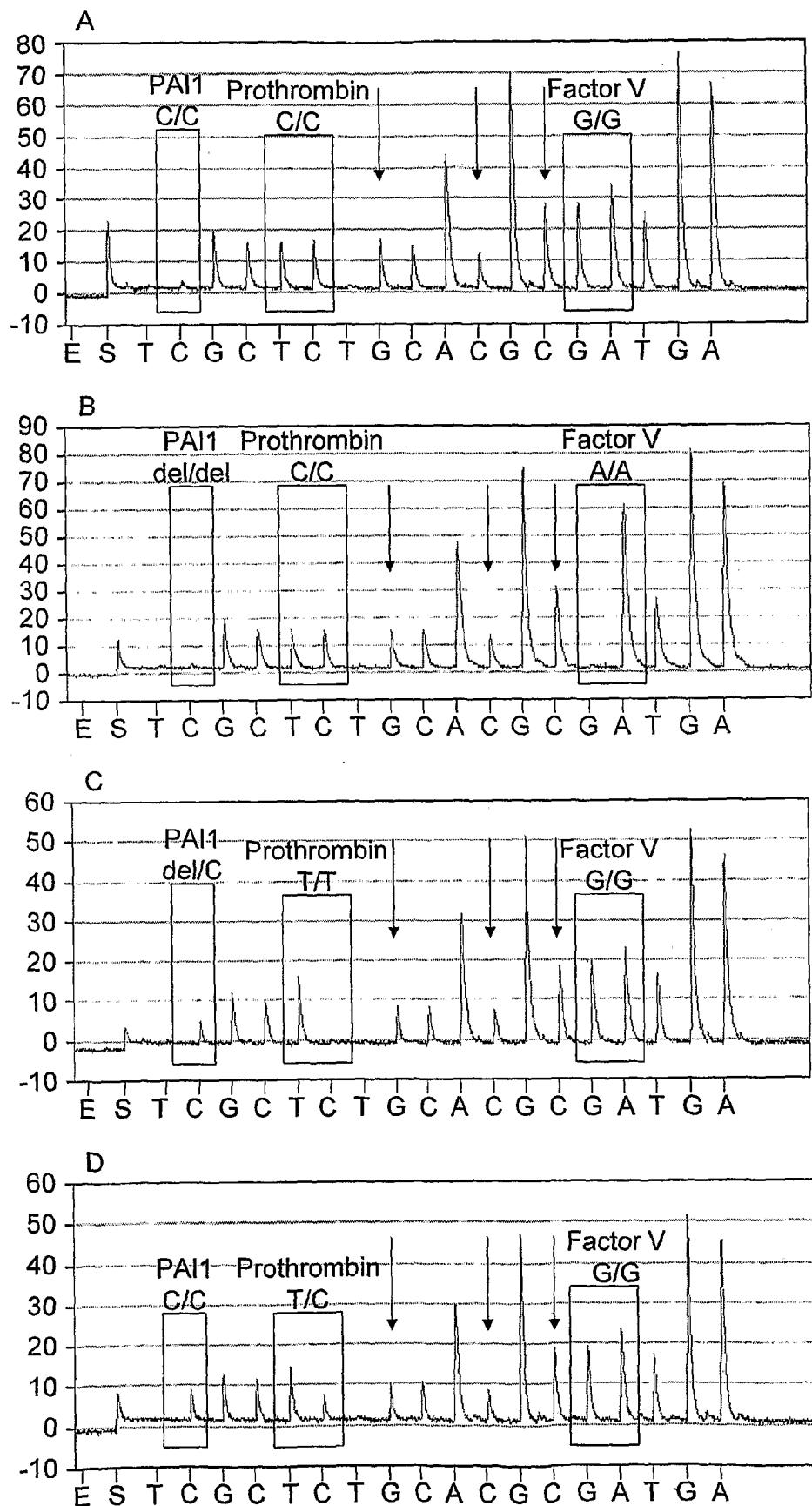


FIG. 14

SUBSTITUTE SHEET (RULE 26)

17 / 19

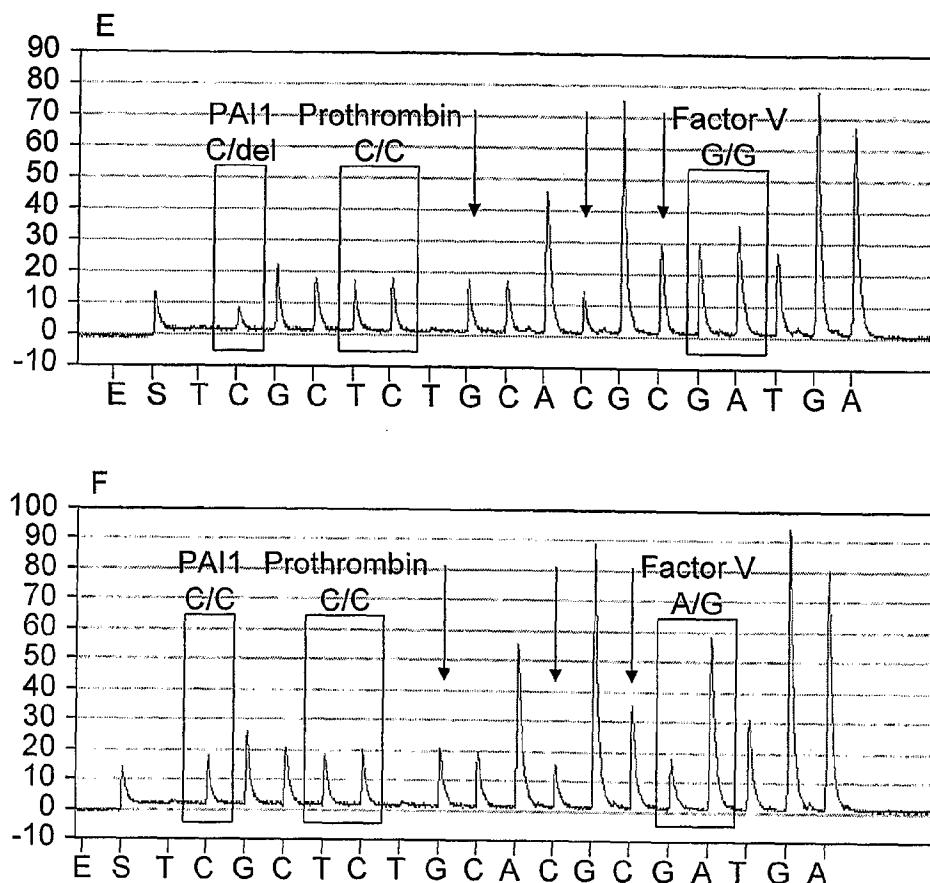


FIG. 14 CONT'D

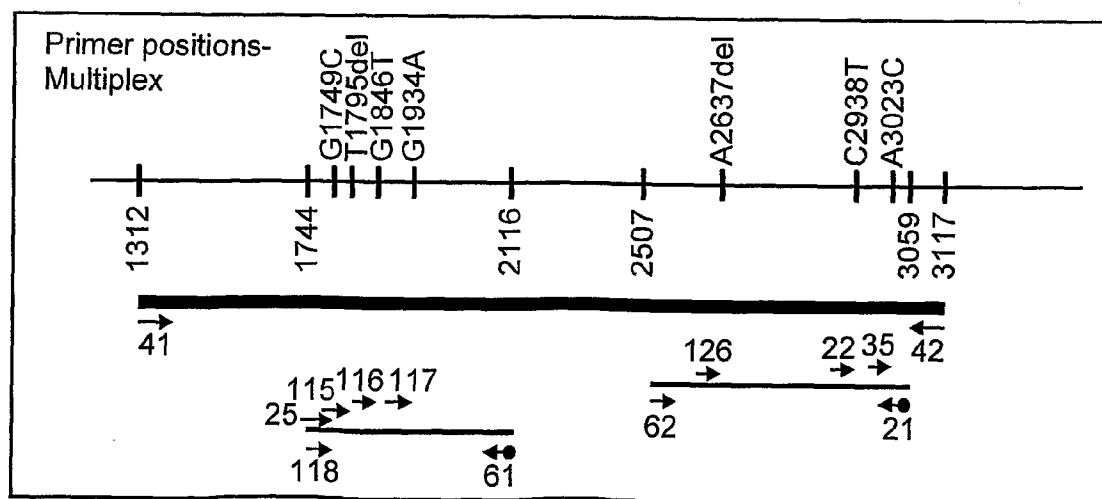


FIG. 15

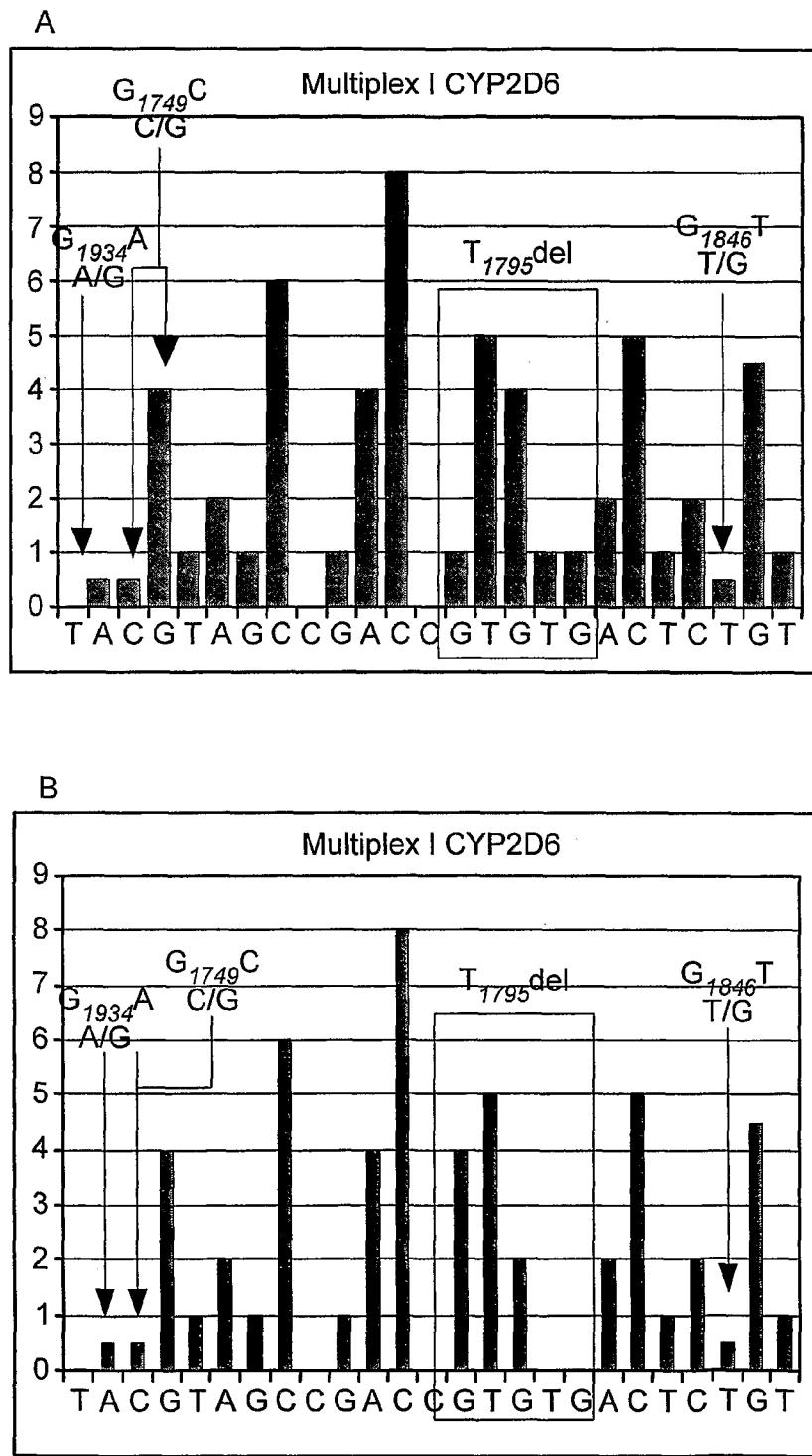


FIG. 16

- G1846T: GCCAACCACTCC G/T GT
- G1934A: G/ AGACGCCCTTCG
- T1795del: GAGCAG(T)GGGTGACCG
- G1749C: G/ CTCCACCTTGC

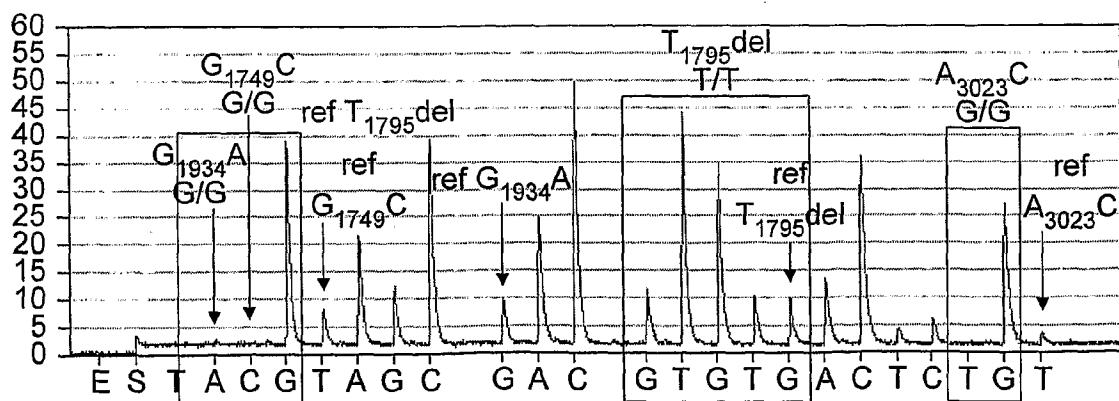
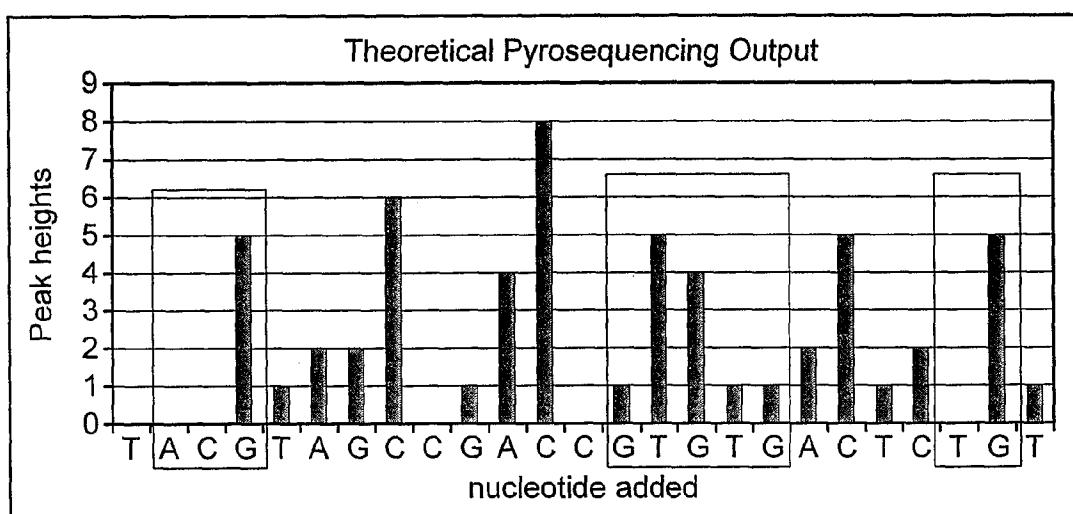


FIG. 17